Peptide targeting by spontaneous isopeptide bond formation



Bijan Zakeri, HBSc, MSc

Department of Biochemistry
and St. Peter's College
University of Oxford

Trinity 2011

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, University of Oxford.

© Copyright by Bijan Zakeri, September 2011

DPhil Thesis, Trinity 2011 Bijan Zakeri
Department of Biochemistry and St. Peter's College

Abstract

Peptide targeting by spontaneous isopeptide bond formation

Peptide fusion tags are fundamental for the identification, detection, and capture of proteins in biological assays. Commonly used peptide fusion tags rely on temporary non-covalent interactions for binding, which can put constraints on assay sensitivity. Here, peptide fusion tags were developed that could specifically interact with protein binding partners via spontaneous and irreversible isopeptide bond formation. To develop covalently interacting peptide-protein pairs, outer-membrane proteins from Gram-positive bacteria that form autocatalyzed intramolecular isopeptide bonds were dissected to generate a short peptide fragment and a protein binding partner. Initially, the major pilin subunit Spy0128 from *Streptococcus pyogenes* was split to develop the 16 residue isopeptag peptide and the 31 kDa pilin-C protein partner. The isopeptag:pilin-C pair were able to react via spontaneous isopeptide bond formation between an Asn residue in isopeptag and a Lys residue in pilin-C without the requirement for any accessory factors, and with a yield of 60% after a 72 hr reaction. Reconstitution between the isopeptag:pilin-C pair was robust and occurred under all biologically relevant conditions tested, and also in the complex environment of a bacterial cytosol and on the surface of mammalian cells. A similar approach was also used to dissect the small CnaB2 domain that is part of the large FbaB fibronectin-binding protein from *S. pyogenes*. This led to the development of a more efficient peptide-protein pair, which was rationally modified to generate the highly optimized SpyTag:SpyCatcher pair. SpyTag is a 13 amino acid peptide with a reactive Asp that forms a spontaneous intermolecular isopeptide bond with a Lys present in the 12 kDa SpyCatcher binding partner. In a reaction with SpyTag, over 40% of SpyCatcher was depleted after 1 min and SpyCatcher could no longer be detected after 2 hr. The SpyTag and SpyCatcher reaction did not require any accessory factors and proceeded efficiently at a range of biologically relevant temperatures, pH values, concentrations, buffer compositions, and in the presence of commonly used detergents. The SpyTag:SpyCatcher technology was also used for specific cell surface labelling on mammalian cell membranes. SpyTag and SpyCatcher are both composed of the regular 20 amino acids and can therefore be genetically encoded as fusion constructs for a variety of *in vitro* and *in vivo* applications. Potential applications of the SpyTag:SpyCatcher technology include specific cell surface labelling, the development of novel protein architectures, and the covalent and irreversible capture of target proteins in biological assays.

Authorship Declaration

I declare that this thesis represents entirely my work, except where acknowledged below. This thesis has not been submitted for any other degree.

Parts of this thesis done by other people:

- Section 4.2.5.2. NMR spectroscopy of ^{15}N-Cna110 and Cna peptide reconstitution.
- A few constructs were cloned by Dr. Mark Howarth.

Publications in the course of my thesis include:

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.

Howarth,M. and Zakeri,B. (2010). Peptide tag systems that spontaneously form an irreversible link to protein partners via isopeptide bonds. *United Kingdom Patent Application No. 1002362.0.*

Acknowledgements

To my supervisor Dr. Mark Howarth, thank you for your support and guidance throughout this project. You taught me the scientific method and demonstrated the dedication and sacrifice that is required to become a great scientist. I would like to thank my funding bodies, the Clarendon Fund and the Department of Biochemistry, for providing me with this great opportunity to study at a world renowned institution. I would also like to thank Dr. Uli Schwarz-Linek for a fruitful collaboration.

I would like to dedicate this work to my family Dr. Forough Farrokhyar, Dr. Mahmood Zakeri, Dr. Arash Zakeri, and Bita Zakeri who is currently a PhD candidate. In particular, I am greatly indebted to my mother who is the source of my inspiration and has always been there to encourage me. I thank you for always being there to support me in my troubles and to share in my joys.

Table of Contents

Abstract ii
Authorship declaration iii
Acknowledgements iv
Table of Contents v
Abbreviations ix
List of Figures and Tables x

Chapter 1: Introduction 1
1.1. Peptide fusion tags 2
1.1.1. Properties of commonly used peptide fusion tags 3
1.1.2. Binding properties of peptides 5
1.2. Covalent linking of polypeptides 6
1.2.1. Disulfide bonds 6
1.2.2. Native chemical ligation 7
1.2.3. Intein mediated post-translational protein splicing 9
1.2.4. Other methods of covalently linking polypeptide sequences 13
1.3. Isopeptide bonds 14
1.3.1. Intermolecular isopeptide bond formation 15
1.3.2. Spontaneous intramolecular isopeptide bond formation 17
1.3.3. Other intramolecular isopeptide bond containing proteins 20
1.3.4. Mechanism of autocatalytic isopeptide bond formation 24
1.4. Gram-positive bacterial pathogens use intramolecular isopeptide bonds to stabilize outer membrane proteins 26
1.4.1. Pilus assembly in Gram-positive bacteria 26
1.4.2. Biological function of intramolecular isopeptide bonds 29
1.5. Protein splitting technology 31
1.6. Objectives of this work 34

Chapter 2: Methods 38
2.1. Cloning 39
2.1.1. General cloning procedures 39
2.1.2. Cloning of Spy0128 constructs 40
2.1.3. Cloning of ACE19 constructs 43
2.1.4. Cloning of CnaB2 constructs 44
2.2. Protein expression and purification 48
2.2.1. General protein expression and purification procedures 48
2.2.2. Expression of ^{15}N-Cna110 50
2.3. SDS-PAGE 50
2.4. Trypsin digestion reactions 51
2.5. Reconstitution reactions 51
2.5.1. Reconstitution of split Spy0128 constructs 51
2.5.2. Reconstitution of split ACE19 constructs 53
2.5.3. Reconstitution of Cna peptide with Cna110 53
2.5.4. Reconstitution of ^{15}N-Cna110 53
2.5.5. Reconstitution of split CnaB2 constructs 54
2.6. Mass spectrometry 55
2.6.1. Analysis of Spy0128 55
2.6.2. Analysis of split Spy0128 constructs 56
2.6.3. Analysis of ACE19 and split CnaB2 constructs 56
2.6.4. Analysis of ^{15}N-Cna110 and Cna peptide 56

2.7. SPASM searches 57
2.8. NMR spectrometry 57
2.9. Mammalian cell culture and cellular imaging 58
2.10. Phage display 60
2.10.1. Tethering of CnaB2 constructs to magnetic beads and ELISA plates 60
2.10.2. Phage display panning experiments 61
2.10.3. Western blots for phage display screens 63

Chapter 3: Dissection of Spy0128 to develop a spontaneously and covalently reconstituting peptide-protein pair 66
3.1. Recombinant Spy0128 was able to form two intramolecular isopeptide bonds 67
3.2. Spy0128 was resistant to trypsin digestion 68
3.3. Splitting of Spy0128 71
3.3.1. Strategy for the development of covalent peptide-protein pairs 71
3.3.2. Splitting of Spy0128 in the N-terminal domain 73
3.3.3. Splitting of Spy0128 in the C-terminal domain 75
3.4. Optimization of split constructs 77
3.4.1. Removal of the N-terminal 6X His tag from CB-MBP 77
3.4.2. Rationale for the generation of the CB variant isopeptag 78
3.4.3. Rationale for optimization of the lengths of isopeptag and pilin-C -7 79
3.4.4. Reconstitution efficiency of the isopeptag and pilin-C -7 variants 81
3.5. Characterization of isopeptag and pilin-C reactivity 82
3.5.1. Isopeptag-MBP and pilin-C specifically reconstituted *in vitro* 82
3.5.2. Mass spectrometric analysis of the isopeptag-MBP and pilin-C reaction 83
3.5.3. Time course of isopeptag-MBP and pilin-C reconstitution 85
3.5.4. Isopeptag-MBP and pilin-C reconstitution at varying ratios 86
3.5.5. Concentration-dependence of reconstitution between isopeptag-MBP and pilin-C 87
3.5.6. Temperature-dependence of isopeptag-MBP and pilin-C reconstitution 88
3.5.7. pH-dependence of isopeptag-MBP and pilin-C reconstitution 89
3.5.8. Isopeptag-MBP and pilin-C reconstitution in various buffers and detergents 90
3.6. Isopeptag-MBP and pilin-C could reconstitute inside a bacterial cell 92
3.7. Isopeptag and pilin-C were used for cell surface labelling 93
3.8. Summary 95

Chapter 4: Investigation of smaller protein scaffolds for the development of an efficiently reconstituting and covalent peptide-protein pair 99
4.1. ACE19 contained an intramolecular isopeptide bond 100
4.1.1. Approach for structural similarity searches for the discovery of previously unrecognized intramolecular isopeptide bond containing proteins 100
4.1.2. Results from structural similarity searches 101
4.1.3. Further analysis of the ACE19 crystal structure 102
4.1.4. Mass spectrometry of ACE19 104
4.1.5. Dissection of ACE19 105
4.1.6. Split ACE19 constructs were able to covalently reconstitute 106
4.2. Investigations of the CnaB2 protein 108
4.2.1. CnaB2 is a small protein domain that contains an intramolecular isopeptide bond 108
4.2.2. Dissection of CnaB2 109
4.2.3. Split CnaB2 constructs were able to covalently reconstitute 111
4.2.4. Preliminary characterization of the split CnaB2 reconstituting pair 114
4.2.5. Preliminary structural studies on split protein reconstitution 116

4.2.5.1. Cna110 and Cna peptide covalently reconstituted 116
4.2.5.2. NMR spectroscopy of ^{15}N-Cna110 and Cna peptide reconstitution 117
4.3. Summary 122

Chapter 5: Phage display and rational modifications to develop an efficiently reconstituting covalent peptide-protein pair 127
5.1. Phage display for isolating efficiently reacting covalent peptides 128
5.1.1. Introduction to phage display 128
5.1.1.1. Bacteriophage M13 phage display libraries 128
5.1.1.2. General phage display panning protocol 129
5.1.2. Phage display screens for efficient and covalently binding peptides: Experimental setup 132
5.1.3. Phage display screens for efficient and covalently binding peptides: Results ... 134
5.2. Rational modification experiments for isolating efficiently reacting covalent peptide-protein pairs 140
5.2.1. Rational modifications of Cna114 and PepA 140
5.2.2. Approaches taken to optimize Cna114 reconstitution efficiency 141
5.2.3. Effects of Cna114 modifications on reconstitution efficiency 143
5.2.4. Optimized Cna114 binding partner 145
5.2.5. Approaches taken to optimize PepA reconstitution efficiency 146
5.2.5.1. Rationale for PepD/E/F/G/H/I/O 146
5.2.5.2. Rationale for PepC/J/K 148
5.2.5.3. Rationale for PepN 148
5.2.5.4. Rationale for PepL/M/P 149
5.2.6. Reconstitution of the PepA mutant series versus SpyCatcher 150
5.2.6.1. PepB/D/E/F/G/H/I/O reconstitution with SpyCatcher 150
5.2.6.2. PepC/J/K reconstitution with SpyCatcher 152
5.2.6.3. PepN reconstitution with SpyCatcher 152
5.2.6.4. PepL/M/P reconstitution with SpyCatcher 153
5.2.6.5. Reconstitution between Cna114 variants and SpyTag 154
5.3. Summary 155

Chapter 6: Characterization of the SpyTag:SpyCatcher technology ... 159
6.1. The SpyTag:SpyCatcher technology 160
6.2. Mass spectrometry of the SpyCatcher and Cna peptide reaction 162
6.3. Speed of reconstitution between SpyTag-MBP and SpyCatcher 163
6.4. SpyTag-MBP and SpyCatcher reconstitution at varying ratios 166
6.5. Concentration-dependence of SpyTag-MBP and SpyCatcher reconstitution 167
6.6. Temperature-dependence of SpyTag-MBP and SpyCatcher reconstitution 168
6.7. pH-dependence of SpyTag-MBP and SpyCatcher reconstitution 170
6.8. SpyTag-MBP and SpyCatcher reconstitution in commonly used buffers 171
6.9. SpyTag-MBP and SpyCatcher reconstitution in the presence of detergent ... 172
6.10. Reversibility test of the SpyTag-MBP:SpyCatcher covalent reaction 174
6.11. Cross-reactivity of the SpyTag:SpyCatcher and isopeptag:pilin-C technologies 176
6.12. SpyTag and SpyCatcher reconstitution on the surface of mammalian cells 178
6.13. Summary 181

Chapter 7: Summary and General Discussion 184
7.1. A general method for developing intermolecular isopeptide bond forming peptide-protein pairs 185
7.1.1. Dissection of Spy0128 186

7.1.2. Dissection of ACE19 187
7.1.3. Dissection of CnaB2 188
7.2. A new class of covalently binding peptide fusion tags 189
7.3. A new class of split proteins 190
7.4. Structural insights in to split protein reconstitution 193
7.5. Insights in to intramolecular isopeptide bond formation 194
7.6. Further enhancing the reconstitution efficiency of SpyTag and SpyCatcher ... 195
7.7. Potential Applications of the SpyTag:SpyCatcher technology 196
7.7.1. Signal amplification of binding events 196
7.7.2. Assembly of novel protein architectures 199
7.7.3. Development of a series of isopeptide bond forming peptide fusion tags 202
7.7.4. The SpyTag:SpyCatcher system might be used as a tool for sensitive .. 203
detection of target proteins
7.7.5. The SpyTag:SpyCatcher system might be used to tether target 204
proteins to a solid support
7.8. Concluding remarks 205

References 206

Appendix 213
Reprint of Zakeri and Howarth, 2010 216

Abbreviations

ACE19	adhesin of collagen from Enterococci (19 kDa)
BHS	biotin holoenzyme synthetase
BSA	bovine serum albumin
CFP	cyan fluorescent protein
DMEM	Dulbecco's modified eagle's medium
DMF	*N,N*-dimethylformamide
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
ELISA	enzyme-linked immunosorbent assay
GdmCl	guanidinium hydrochloride
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
HSQC	heteronuclear single quantum coherence
IMAC	immobilized metal-affinity chromatography
IPTG	isopropyl β-D-thiogalactoside
K_d	dissociation constant
LB	Luria-Bertani
MBP	maltose binding protein
NHS	*N*-hydroxysuccinimide
NMR	nuclear magnetic resonance
NP40	Nonidet P-40
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
pfu	plaque forming units
PMSF	phenylmethylsulfonyl fluoride
SDS	sodium dodecyl sulphate
SPASM	SPatial Arrangements of Side chains and Main chains
TBS	tris buffered saline
TEV	Tobacco Etch Virus
TM	trans-membrane helix
Xgal	5-bromo-4-chloro-3-indolyl β-D-galactoside

List of Figures and Tables

Chapter 1: Introduction

Figures:

1.1. Peptide fusion tags allow the capture of proteins of interest for various applications 3
1.2. Disulfide bond formation 7
1.3. Native chemical ligation 9
1.4. Intein mediated splicing 12
1.5. Isopeptide bond formation 16
1.6. Spy0128 contains two spontaneously formed intramolecular isopeptide bonds 18
1.7. Linear covalent connectivity of Spy0128 18
1.8. Mechanism of autocatalytic isopeptide bond formation 25
1.9. A schematic representation of sortase mediated assembly of Gram-positive bacterial pili 28
1.10. Splitting of a hypothetical fluorescent protein 33

Tables:

1.1. A survey of experimentally confirmed intramolecular isopeptide bond containing proteins 22

Chapter 3: Dissection of Spy0128 to develop a spontaneously and covalently reconstituting peptide-protein pair

Figures:

3.1. Recombinant Spy0128 formed two intramolecular isopeptide bonds 68
3.2. Trypsin digestion of Spy0128 and its isopeptide-deficient mutants 71
3.3. Strategic locations to split Spy0128 73
3.4. Spontaneous intermolecular isopeptide bond formation between MBP-isopeptag-N and pilin-N 74
3.5. Spontaneous intermolecular isopeptide bond formation between CB-MBP and pilin-C -7 76
3.6. Reconstitution between CBΔHis-MBP and pilin-C -7 78
3.7. Phe^{307} is a hydrophobic residue in an unstructured region 79
3.8. Optimization of the lengths of isopeptag and pilin-C -7 80
3.9. Covalent reconstitution between isopeptag-MBP and pilin-C 82
3.10. Mass spectrometric analysis of isopeptag-MBP and pilin-C reconstitution . 84
3.11. Time course of isopeptide bond formation 86
3.12. Reconstitution of one reactant against varying concentrations of another .. 87
3.13. Concentration-dependence of reconstitution between isopeptag-MBP and pilin-C 88
3.14. Temperature-dependence of isopeptag-MBP and pilin-C reconstitution 89
3.15. pH-dependence of isopeptag-MBP and pilin-C reconstitution 90
3.16. Isopeptag-MBP and pilin-C reconstitution in various buffers and detergents 91
3.17. Isopeptag-MBP and pilin-C could reconstitute inside a bacterial cell 93
3.18. Cell surface labelling with isopeptag and pilin-C 95

Chapter 4: Investigation of smaller protein scaffolds for the development of an efficiently reconstituting and covalent peptide-protein pair

Figures:
4.1. Isopeptide bond forming residues used as templates for SPASM structural similarity searches 102
4.2. ACE19 had one potential intramolecular isopeptide bond 103
4.3. Comparison of three intramolecular isopeptide bond containing proteins 104
4.4. Mass spectrometry of ACE19 105
4.5. Split ACE19 constructs could covalently reconstitute 107
4.6. CnaB2 contains an intramolecular isopeptide bond 109
4.7. Splitting of CnaB2 111
4.8. Reconstitution of split CnaB2 constructs 113
4.9. Time course of PepA-MBP and Cna110 covalent reconstitution 115
4.10. Temperature-dependence of PepA-MBP and Cna110 reconstitution 115
4.11. Preparation of samples for structural studies on split protein reconstitution 120
4.12. Analysis of ^{15}N-Cna110 and Cna peptide reconstitution with NMR spectroscopy and mass spectrometry 121

Chapter 5: Phage display and rational modifications to develop an efficiently reconstituting covalent peptide-protein pair

Figures:
5.1. Phage display 131
5.2. Phage display approaches 139
5.3. Optimization of the Cna114 platform 142
5.4. A close view of the C-terminal β-strand in CnaB2 147
5.5. Generation of PepL, PepM, and PepP 149
5.6. Reconstitution between SpyCatcher and PepA variants 151
5.7. Reconstitution between SpyTag-MBP and Cna114 variants 154

Tables:
5.1. A summary of phage display experiments 137
5.2. Summary of the results obtained from a series of mutations made to Cna114 145
5.3. A summary of the PepA variants and their activities 147

Chapter 6: Characterization of the SpyTag:SpyCatcher technology

Figures:
6.1. Covalent reconstitution between SpyTag-MBP and SpyCatcher 161
6.2. Mass spectrometry of the SpyCatcher and Cna peptide reaction 163
6.3. Speed of reconstitution between SpyTag-MBP and SpyCatcher 165
6.4. SpyTag-MBP and SpyCatcher reconstitution at varying protein ratios 167
6.5. Concentration-dependence of SpyTag-MBP and SpyCatcher reconstitution 168
6.6. Temperature-dependence of SpyTag-MBP and SpyCatcher covalent reconstitution 169
6.7. SpyTag-MBP and SpyCatcher reconstitution proceeded at a range of pH values 171
6.8. SpyTag-MBP and SpyCatcher reconstitution in a range of biological buffers 172
6.9. SpyTag-MBP and SpyCatcher reconstituted in the presence of commonly used detergents 173
6.10. SpyTag-MBP and SpyCatcher reconstitution was not reversible 175
6.11. SpyTag:SpyCatcher and isopeptag:pilin-C technologies could cross-react 177
6.12. SpyTag could be specifically labelled on mammalian cells 180

Chapter 7: Summary and General Discussion

Figures:
7.1. Split protein technologies 192
7.2. Self-assembling of SpyTag-SpyCatcher fusions 198
7.3. Potential applications of SpyTag and SpyCatcher for the assembly 201
of novel protein architectures

Appendix

Figures:
A.1. Reconstitution of Cna114 mutants 214
A.2. Reconstitution of Cna114 variants 215

Chapter 1:

Introduction

Chapter 1: Introduction

1.1. Peptide fusion tags

Fundamental to biochemical research is the ability to isolate or target a protein of interest (Terpe, 2003). A simple, fast, and economical method of doing so would be to fuse a peptide tag to a protein of interest, thereby providing a means of targeting or capturing the fusion protein. Peptide fusion tags are vital to many biological applications including protein purification, protein identification (e.g. Western blot analysis, cellular imaging), and immobilization of a protein to a solid support (Fig. 1.1) (Terpe, 2003; Xu and Evans, Jr., 2005; Popp et al., 2007). An ideal peptide tag would have the following properties: (i) it would be small, ideally 5-15 amino acids to aid in its introduction in to desired genes and to limit the introduction of foreign sequences in the protein of interest, (ii) it would not interfere with protein folding or function (Terpe, 2003), (iii) it would be soluble, so that it can be displayed in solution and interact with its binding partner, (iv) it would have high specificity for its binding partner to avoid non-specific interactions, (v) it would have high affinity for its binding partner to allow for efficient binding, (vi) it would retain function when bound to the N-terminal, C-terminal, or internal loop regions of a protein, and (vii) it would have a robust interaction with its binding partner so that it would still bind efficiently in a range of temperatures and pH values, and regardless of the redox state, buffer composition, and the presence of detergents (Terpe, 2003). The perfect peptide tag does not exist; however those that are currently available each possess several of the above characteristics and one must choose the peptide tag that best meets the experimental requirements (Terpe, 2003).

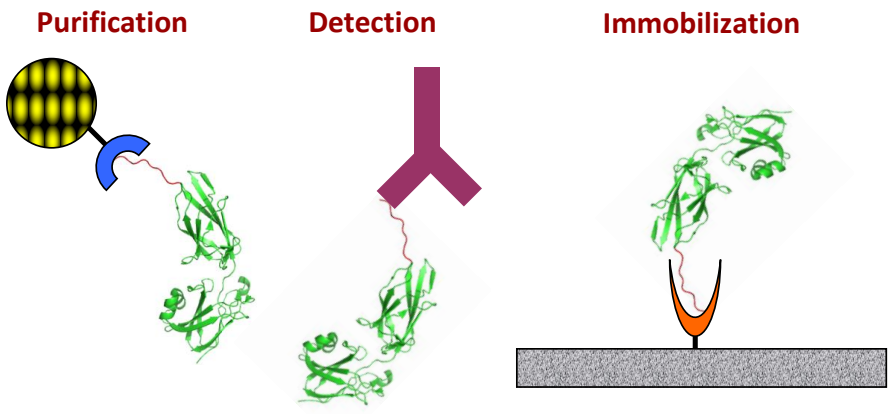


Figure 1.1. Peptide fusion tags allow the capture of proteins of interest for various applications. The figures depict a target protein (green) fused to a peptide tag (red). Left: tagged proteins can be purified from complex solutions by binding of peptide tags to, for example the binding partner coupled to agarose beads, where non-binding proteins can be washed away and the purified protein subsequently eluted. Middle: peptide tags can be detected using antibodies for applications such as Western blot analysis. Right: proteins can be immobilized on to a solid support via their peptide tags for applications such as compound library screenings.

1.1.1. Properties of commonly used peptide fusion tags

Although there are a large number of peptide fusion tags available for various applications, only a few are commonly used. For protein purification the His tag is routinely employed and it consists of 2-10 (most commonly 6) consecutive histidine residues that allow for protein purification via immobilized metal-affinity chromatography (IMAC) (Terpe, 2003). The imidazole rings in the histidine side chains interact strongly with metal ions immobilized on a solid surface such as agarose beads, thereby allowing specific capture of His tag containing proteins from a complex mixture such as a cell lysate (Terpe, 2003; Qiagen, 2002). His tags can also be utilized for the detection of fusion proteins via an anti-His antibody that specifically binds His tags with a typical dissociation constant (K_d) of 5 x 10^{-8} M (Qiagen, 2002). Another commonly used peptide fusion tag is the 11 amino acid c-myc tag that can be targeted with the 9E10 antibody for applications such as detection of fusion proteins in Western blot analysis or purification of fusion constructs via immunoprecipitation (Terpe, 2003). The interaction between the c-

myc tag and the 9E10 antibody has a K_d of 5 x 10^{-7} M and is therefore weaker than the interaction between the His tag and its corresponding antibody (Krauss et al., 2008).

Innovative peptide tagging systems have also been developed based on the high affinity interaction between streptavidin and its ligand biotin (K_d 4 x 10^{-14} M (Green, 1990)). In one approach, a 10 residue peptide fusion tag has been developed named Strep-tag II that binds Strep-Tactin (a streptavidin variant optimized to bind Strep-tag II) and can be utilized under stringent conditions such as in the presence of high salt, detergent, chelating agents, and reducing agents (Schmidt et al., 1996; Terpe, 2003; Voss and Skerra, 1997). Strep-tag II is a peptide that binds Strep-Tactin in the same binding pocket that is occupied by its natural ligand biotin, however at a much lower affinity with a K_d of 1 x 10^{-6} M (Schmidt et al., 1996; Voss and Skerra, 1997). The weak Strep-tag II and Strep-Tactin interaction allows the capture of Strep-tag II fusion constructs with immobilized Strep-Tactin followed by efficient elution using biotin variants (Voss and Skerra, 1997). In general, most commonly used peptide fusion tags interact with their binding partners through weak non-covalent interactions leading to high dissociation constants. In an innovative approach to develop a high affinity peptide fusion tag, the high affinity streptavidin-biotin interaction was adapted to develop the AviTag system (Beckett et al., 1999; Cull and Schatz, 2000). The biotinylation of certain enzymes is a post-translational modification that is performed by the biotin holoenzyme synthetase (BHS) class of enzymes, where a biotin molecule is covalently linked to a conserved Lys residue in an ATP dependent reaction (Beckett et al., 1999). To adapt this process for the development of a peptide fusion tag, the 15 residue AviTag peptide was selected whose sequence was optimized to be recognized and efficiently biotinylated by the enzyme BirA (a BHS enzyme) (Beckett et al., 1999). Avitag can be fused to a protein of interest and biotinylated using BirA both *in vitro* and *in vivo*, where the biotinylated AviTag fusion construct can subsequently be captured by streptavidin for various applications such as detection, immobilization, or purification of target proteins (Beckett et al., 1999; Cull and Schatz, 2000; Howarth and Ting, 2008).

1.1.2. Binding properties of peptides

The affinity and specificity of peptide fusion tags for their binding partners are vital parameters for biological assays, which are defined by peptide-protein interactions. These interactions, whether between protein-ligand or antibody-antigen complexes, are temporary non-covalent interactions (Houk et al., 2003). In biological assays, it would be highly desirable to have the peptide tag-binding partner interactions to be much stronger than the interactions under study. As such, the binding affinity of the peptide fusion tags would not be a limiting factor in investigations and stringent experimental conditions can be used as required. However, the binding affinities of commonly used peptide fusion tags fall within the same range as those of antibody-antigen and enzyme-inhibitor binding affinities, notably having K_d values in the micromolar to nanomolar range (Houk et al., 2003). The only class of biological interactions that have stronger binding affinities (K_d values in the femtomolar range) are enzyme-transition state complexes, which possess such strong binding due to strong electrostatic and covalent interactions, and conformationally favourable binding cavities (Houk et al., 2003). Protein-protein interactions are facilitated by the hydrophobic effect, hydrogen bonding, and electrostatic interactions, however in peptide-protein complexes these interactions are limited due to the small surface area available for binding as a result of the limited size of the peptides (Houk et al., 2003). Houk et al. have found that there was a clear correlation between the surface area of a binding interface and the binding affinity of an interaction (Houk et al., 2003). It is difficult for proteins to forge strong interactions with peptides due to their limited surface area and structural flexibility. Therefore, there is an inherent limitation to the binding strength that can be achieved by conventional peptide-protein interactions (Houk et al., 2003). *In vitro* evolution techniques can be used in an attempt to increase binding affinities between peptide-protein pairs. However, as long as binding is mediated through non-covalent interactions, these techniques are unlikely to achieve significant enhancements in binding strengths that can rival covalent bonds. Hence, to develop peptide-protein interacting pairs that can achieve maximum affinity, we need to look

beyond conventional molecular interactions. We need to develop peptide-protein pairs that can interact covalently to overcome the restraints imposed by the limited surface area of peptides and weak binding interactions.

1.2. Covalent linking of polypeptides

To achieve maximum affinity between peptide-protein complexes, we must look to different methods of covalently coupling two polypeptide sequences.

1.2.1. Disulfide bonds

There are many ways of linking polypeptides through covalent bonds, the simplest method being through disulfide bond formation between two cysteine residues, one present in each partner. Cysteines possess reactive thiol groups that when brought in to close proximity readily oxidize to form cystines (Fig. 1.2). Naturally, proteins use disulfide bonds as a method of introducing enhanced structural stability (Voet and Voet, 2004). Disulfide bond formation does not generally occur in the cytoplasm because reductants such as thioredoxin and glutathione function to reduce disulfide linkages in intracellular proteins (Meyer et al., 2009; Voet and Voet, 2004). Since thiol moieties are very reactive and proteins can possess many cysteine residues leading to multiple disulfide bonds being formed, there are mechanisms to ensure that the correct disulfide bridges are formed between specific cysteine residues, otherwise proteins would mis-fold (Voet and Voet, 2004). In eukaryotes, disulfide bond formation is mediated by protein disulfide isomerase in the endoplasmic reticulum, which has an oxidizing environment, to ensure the correct folding of disulfide bonded proteins (Voet and Voet, 2004). In bacteria, disulfide bonded proteins are far more common in Gram-negative organisms due to a well established oxidizing system. Oxidation occurs in the periplasm, where the enzyme DsbA mediates oxidative disulfide bond formation (Kadokura and Beckwith, 2010). The structural stability provided by disulfide bonds help proteins to overcome the harsh

extracellular environments that are prone to temperature and pH fluctuations (Voet and Voet, 2004).

Disulfide bond formation is therefore a simple method by which we could covalently link polypeptides. However, it would not be ideal as a mechanism of developing stable and covalent peptide-protein complexes because of the complications that would be introduced. Reactive thiol groups in peptides/proteins could lead to undesired homodimerization during purification steps. This would lead to the additional steps of using reducing agents to reduce disulfide bridges and the subsequent removal of the reducing agents, which would likely result in a decrease in yield of the desired monomeric peptide/protein. Furthermore, peptide-protein reconstitution reactions would be prone to homodimerization side-reactions and non-specific disulfide bond formation between the peptide/protein and other proteins containing Cys residues.

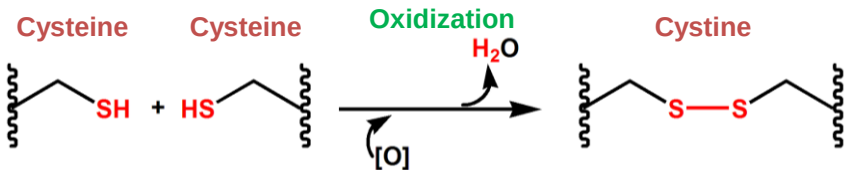


Figure 1.2. Disulfide bond formation. Two cysteine residues are oxidized to form a cystine. Adapted from (Voet and Voet, 2004).

1.2.2. Native chemical ligation

A more desirable method of linking polypeptides would be through a peptide (also known as amide) bond, thereby creating a more stable polypeptide complex. Nature's solution is to use enzymatic catalysis, namely ribosomal polypeptide synthesis, which is difficult and inefficient to emulate *in vitro*. As a more feasible alternative, chemical synthetic approaches to polypeptide synthesis were developed (Dawson and Kent, 2000). With this technique, chemists sought to synthetically construct biologically active proteins (Dawson and Kent, 2000). However, a major drawback of this technique was that the

synthesized peptides were limited to ~50 amino acids (Dawson and Kent, 2000). To address this issue native chemical ligation was developed, where two unprotected peptides could be linked through an amide bond under denaturing conditions and in aqueous solution at neutral pH (Dawson and Kent, 2000). With this approach, small peptides could be made separately and linked together through a peptide bond to construct a whole protein. The reaction proceeded through thioester-mediated chemoselective coupling, where one peptide contained a C-terminal thioester and the other an N-terminal cysteine residue (Fig. 1.3) (Wilken and Kent, 1998). However, this technique has several limitations including that a cysteine residue is left behind at the coupling junction in the final product, reactions must occur under denaturing conditions, and it is primarily for linking small peptide fragments.

To overcome some of these limitations, Beligere and Dawson developed conformationally assisted protein ligation (Beligere and Dawson, 1999). In this approach, they developed two peptides with sequences corresponding to that of the full length protein (chymotrypsin inhibitor 2 which is composed of 64 amino acids). They modified the original sequences to accommodate the C-terminal thioester and N-terminal cysteine requirements of native chemical ligation (Beligere and Dawson, 1999). When the peptides were mixed under denaturing conditions, they reacted to form a native peptide bond and produce the full length protein and the reaction took 8 hr to reach completion (Beligere and Dawson, 1999). However, when the peptides were mixed under aqueous conditions, only 30 min was required to achieve complete reaction (Beligere and Dawson, 1999). This rate enhancement was due to the fact that when the peptides were mixed under aqueous conditions, they would self-associate and fold to the proper protein structure (Beligere and Dawson, 1999). This would bring the reactive N- and C-terminal fragments in close proximity, thereby conformationally assisting the reaction (Beligere and Dawson, 1999). They further went on to demonstrate that due to the advantages of conformational assistance, the N-terminal cysteine residue was no longer needed. Changing the N-terminal cysteine to the original methionine still yielded a reactive peptide reaching

complete ligation in 2 hr (Beligere and Dawson, 1999). Although this was a very elegant approach, it had several limitations including that the reactions in the study were conducted at millimolar peptide concentrations, which was not realistic for many biological applications. Also, the reaction required the presence of a terminal thioester moiety (Wilken and Kent, 1998) which meant that one of the binding partners could not be genetically encoded. Moreover, thioesters are highly reactive and would readily react non-specifically with other thiols or amines that are ever present in biological environments. Native chemical ligation is a great approach for the development of synthetic proteins, the introduction of unnatural amino acids in to protein sequences, and the production of proteins that cannot be made in large quantities through recombinant techniques such as toxic proteins or insoluble proteins (Dawson and Kent, 2000). However, for cellular applications the experimental conditions required for native chemical ligation may be limiting.

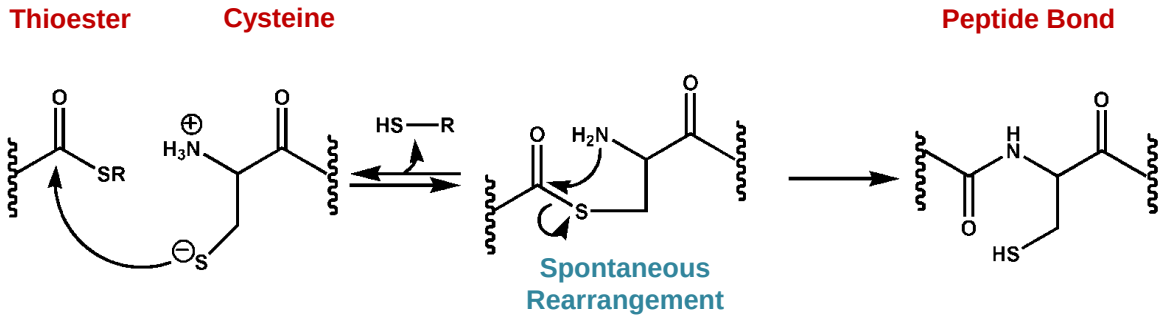


Figure 1.3. Native chemical ligation. A C-terminal thioester and a N-terminal cysteine residue undergo chemoselective coupling followed by a spontaneous rearrangement to give a native peptide bond. Adapted from (Wilken and Kent, 1998).

1.2.3. Intein mediated post-translational protein splicing

Post-translational protein splicing represents another method by which two polypeptide sequences can be covalently linked through peptide bond formation. Some protein precursors have within their polypeptide sequences a non-coding region termed

an intein (Xu and Evans, Jr., 2005). During protein splicing, through a self-catalyzed reaction inteins are removed and the flanking polypeptide fragments, termed exteins, are subsequently joined through a peptide bond to from an active protein (Fig. 1.4a) (Xu and Evans, Jr., 2005). For this reaction to occur the intein must first fold in to the correct conformation, which brings together the ends of the exteins to be spliced (Saleh and Perler, 2006). Subsequently, four nucleophilic reactions occur to excise the intein sequence and ligate the exteins without the requirement of any accessory factors (Fig. 1.4b) (Saleh and Perler, 2006). Generally, for catalysis to occur the N-extein must have a Cys or a Ser at its C-terminal end, the intein must have an Asn at its C-terminal end, and the C-extein must have a good nucleophile such as a Cys, Ser, or Thr at its N-terminal end, although examples with variations of these residue do exist (Saleh and Perler, 2006). The first step in the splicing reaction generates a thioester intermediate through a N-S acyl shift of the N-extein Cys (Fig. 1.4b) (Saleh and Perler, 2006). This is followed by a nucleophilic attack of the thioester with a Cys residue from the C-extein (Fig. 1.4b) (Saleh and Perler, 2006). Subsequent Asn cyclization releases the intein and a S-N acyl shift of the coupled exteins generates a native peptide bond between the ligated exteins (Fig. 1.4b) (Saleh and Perler, 2006; Xu and Evans, Jr., 2005).

Inteins are an elegant method of linking polypeptide chains via a peptide bond. However, if the fragments to be ligated must be part of a single large polypeptide sequence consisting of the intein, then this method would not be application for linking two separate polypeptide sequences for biotechnological applications. Remarkably, there exists naturally split inteins where the intein coding sequence is genetically split and when expressed, the split fragments come together and reconstitute to form a catalytically active intein which can catalyze *trans* splicing (Saleh and Perler, 2006). This natural phenomenon has been exploited to develop biotechnological tools based on both naturally and artificially split inteins, that allow the linking of heterologous exteins in *trans* to produce biologically active proteins (Saleh and Perler, 2006; Lockless and Muir, 2009). In this approach, two polypeptides to be ligated can be fused to split intein fragments,

which can spontaneously reconstitute and catalyze the ligation of the two polypeptides (Saleh and Perler, 2006). Although this is an elegant approach for the ligation of polypeptides and has demonstrated biological applicability, it still has some limitations that restrict its use for certain applications. Due to the reactive thioester intermediates in the splicing reaction there are many side reactions that can occur leading to undesired products and non-specific reactions (Saleh and Perler, 2006). Furthermore, the requirement of catalytically active cysteine residues can make the protein splicing reaction sensitive to the redox status of the environment. Regardless of these concerns inteins have great potential as peptide fusion tags for biotechnological applications. For example, Appleby et al. have reduced the size of split inteins from ~40 amino acids for the smaller reconstituting partner to 6 amino acids (Appleby et al., 2009). This new split intein can be used as a peptide fusion tag for directing polypeptide ligation via *trans* splicing (Appleby et al., 2009).

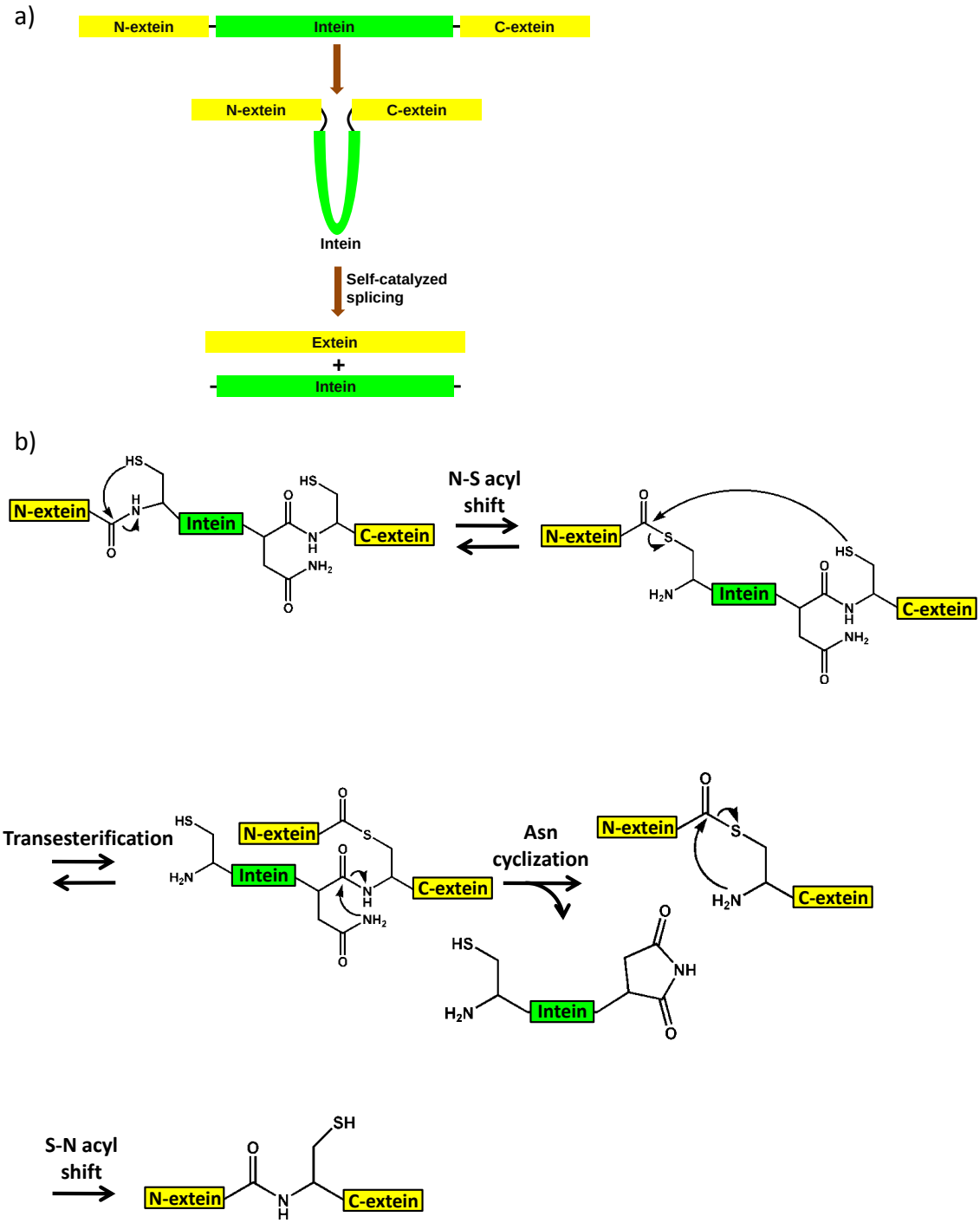


Figure 1.4. Intein mediated splicing. a) The intein domain folds to bring the N- and C-exteins in close proximity and mediates self-splicing to ligate the two extein fragments. b) Mechanism of intein splicing. A thioester intermediate is generated via N-S acyl shift of the N-extein Cys, which then gets nucleophilically attacked by the C-extein Cys. Subsequent Asn cyclization releases the inteins and a S-N acyl shift generates a peptide bond between the two extein fragments. These figures were adapted from (Saleh and Perler, 2006).

1.2.4. Other methods of covalently linking polypeptide sequences

Another innovative technique of mediating peptide bond formation is through peptide ligation using proteases. Rather than catalyzing the cleavage of peptide bonds, proteases have been engineered to catalyze the reverse reaction. For example, two mutations were introduced in active site residues of the serine protease subtilisin to create subtiligase (Abrahmsen et al., 1991). The engineered subtiligase enzyme is able to efficiently catalyze the formation peptide bonds in aqueous solution rather than their cleavage (Abrahmsen et al., 1991; Chang et al., 1994). This technology has been adapted for the tagging of free N-terminal ends of proteins in proteomic studies of post-translational modifications (Atwell and Wells, 1999; Yoshihara et al., 2008). However, a limitation of subtiligase mediated polypeptide ligation is that the C-terminal end of the polypeptide to be ligated must be activated by the addition of a glycolate ester to achieve efficient ligation (Abrahmsen et al., 1991; Braisted et al., 1997). As such, one of the polypeptides to be ligated cannot be genetically encoded and must be synthesized chemically. Furthermore, there is almost no sequence specificity of the N-terminal end of the polypeptide to be ligated (Abrahmsen et al., 1991), which can result in non-specific protein ligation. Accordingly, subtiligase mediated polypeptide ligation would be advantageous for applications such as total protein synthesis (Braisted et al., 1997), introduction of unnatural amino acids in to protein sequences (Braisted et al., 1997), and global tagging of proteins with free N-terminal ends for proteomic studies (Yoshihara et al., 2008).

Alternatively, highly specific tagging of target proteins can be achieved by a technique known as sortagging (Popp et al., 2007). Sortases are a class of enzymes that are used by Gram-positive bacteria to covalently link outer-membrane proteins to the cell wall or to covalently polymerize pilin subunits during pilin assembly (Telford et al., 2006; Mandlik et al., 2008). They do so via recognition of a LPXTG motif on the target protein, which they cleave between the Thr and Gly residues and link the carboxy terminal of the Thr to either the amino group of a pentaglycine precursor of the cell wall or a conserved

lysine residue on a neighbouring pilin subunit (Hendrickx et al., 2011; Telford et al., 2006; Popp et al., 2007). This reaction was adapted for the development of a covalent peptide tagging system. To covalently link two polypeptide sequences, the LPXTG motif could be attached to one polypeptide sequence and the pentaglycine sequence could be attached to the other polypeptide sequence. In the presence of a sortase enzyme, the two polypeptide sequences would be covalently linked together via a sortase catalyzed reaction (Popp et al., 2007; Yamamoto and Nagamune, 2009). This technology has been used in cellular imaging applications where highly specific cell surface labelling has been achieved (Popp et al., 2007; Yamamoto and Nagamune, 2009). In such an application, a cell surface protein was fused to the LPXTG motif allowing for its surface display, and subsequently exogenous sortase enzyme and the pentaglycine motif fused to a signal molecule (a fluorophore or various chemical probes) was added to the cells to achieve specific labelling of cell surface proteins (Popp et al., 2007; Yamamoto and Nagamune, 2009).

The various innovative techniques that have been developed to covalently link polypeptides each have advantages for certain experimental conditions and collectively they provide scientists with a variety of molecular tools. However, it would be desirable to expand the current list of available molecular tools to provide scientists with more options with which to tackle scientific problems.

1.3. Isopeptide bonds

Amide bonds are chemically stable functional groups that link amino acids within a polypeptide chain. However, an amide bond that forms between amino acids outside of the main chain carbons is referred to as an isopeptide bond (Fig. 1.5) (Kang and Baker, 2011).

1.3.1. Intermolecular isopeptide bond formation

Isopeptide bonds are commonly utilized in a variety of biological functions as a means of linking proteins. In eukaryotes, proteins are labelled for downstream processes such as degradation, trafficking, endocystosis, and modulation of expression levels by the strategic attachment of various numbers of ubiquitin, a 76 residue protein (Pickart and Fushman, 2004). To covalently attach ubiquitin to a target protein, an ubiquitin-protein ligase covalently links a Lys side chain on the target protein with the C-terminal Gly residue of ubiquitin via an isopeptide bond (Pickart and Fushman, 2004). Some prokaryotes use a similar method of targeting proteins for degradation, a process termed pupylation (Striebel et al., 2009). In this process, a protein called Pup instead of ubiquitin is attached to target proteins through enzyme catalyzed isopeptide bond formation (Striebel et al., 2009). Isopeptide bonds are also used as a method of strengthening tissues: transglutaminases are a class of enzymes that aid in blood clotting, wound healing, and vascular remodelling (Bakker et al., 2008). Transglutaminases function by cross-linking a Lys side chain in one protein to a Gln side chain of another protein via an isopeptide bond, which provides enhanced mechanical stability to the resulting cross-linked protein complexes (Bakker et al., 2008). These examples of isopeptide bond formation are all mediated through enzyme catalysis; however isopeptide bonds can also be formed spontaneously.

Spontaneous isopeptide bond formation was first observed in the bacteriophage HK97, where it was demonstrated that the HK97 capsid was assembled by the polymerization of a subunit that was covalently cross-linked (Popa et al., 1991; Duda et al., 1995). Subunit cross-linking occurred as a result of autocatalyzed isopeptide bond formation between Lys and Asn side chains in neighbouring subunits (Popa et al., 1991; Duda et al., 1995). These subunits formed covalently assembled 5 and 6 member rings that catenated during capsid assembly to form a structure referred to as protein chainmail (Duda, 1998). In 2000, Wikoff et al. reported the crystal structure of the HK97 capsid that demonstrated the chemical nature of the autocatalyzed isopeptide bonds that cross-linked

the capsid subunits and imparted incredible structural stability to the final capsid structure (Wikoff et al., 2000). The HK97 capsid was composed of 420 subunits that came together to form an icosahedral particle that housed the bacteriophage DNA (Wikoff et al., 2000). During the final step of capsid maturation, each subunit was linked to two neighbouring subunits through isopeptide bonds. Each isopeptide bond formed between Lys and Asn side chains from neighbouring subunits that were positioned in the correct orientation close to a specific Glu residue from a third subunit that catalyzed the reaction (Wikoff et al., 2000). When the Glu residue was mutated, isopeptide bond formation no longer occurred (Wikoff et al., 2000).

These examples demonstrate that isopeptide bonds are used in nature as a method to cross-link and provide structural stability to multi-protein structures. The fact that they are found in eukaryotes, bacteria, and bacteriophage attests to their versatility as a natural molecular tool for linking polypeptides. However, all of the isopeptide bonded proteins that were discussed above were based on intermolecular interactions that either required enzymatic catalysis or the coordinated action of many copies of a single protein. In order to exploit this unique feature of amino acid chemistry for designing molecular tools such as peptide tags, we would require a method of forming an isopeptide bond spontaneously in a bimolecular reaction without the need for any accessory factors.

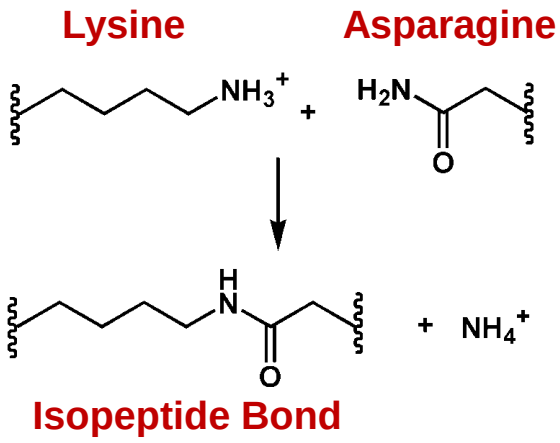


Figure 1.5. Isopeptide bond formation. Isopeptide bond formation between Lys and Asn side chains is illustrated. This figure was adapted from (Zakeri and Howarth, 2010).

1.3.2. Spontaneous intramolecular isopeptide bond formation

Recently, Kang et al. discovered a surprising phenomenon that imparts incredible structural stability to an outer membrane protein of the human pathogen *Streptococcus pyogenes* (Kang et al., 2007). While determining the crystal structure of the protein Spy0128, they discovered that it contained two spontaneously formed intramolecular isopeptide bonds (Fig. 1.6) (Kang et al., 2007). Spy0128 is a 32 kDa protein with an elongated shape and it contains two domains (N- and C-terminal domains) each composed of all β-sheets with a structure that is a variation of the immunoglobulin fold (Kang et al., 2007). Once the protein has folded following translation, the three residues in each domain (Lys, Asn, and Glu) that are involved in isopeptide bond formation come together in a specific orientation and spontaneously react to form an isopeptide bond (Kang et al., 2007). In addition to the orientation of these residues, their surrounding environment is also critical for isopeptide bond formation. These residues are surrounded by a cluster of hydrophobic amino acids that alter the pK_a of the Glu and Lys to make one protonated (Glu) and the other unprotonated (Lys) (Kang et al., 2007). This facilitates a nucleophilic attack by the Lys ε-amino group on the Cγ of Asn, with the intermediate being stabilized by Glu through hydrogen bonding (Kang et al., 2007). The isopeptide bonds are positioned in the folded protein structure to provide an extra dimension of covalent connectivity. In each domain they connect the first and last β-sheet (Fig. 1.6), which in essence makes each domain a folded cyclic polypeptide. This arrangement of isopeptide bonding creates a straight path of covalent connectivity that runs from the N-terminal end to the C-terminal end, thereby creating a covalent 'backbone' for the protein (Fig. 1.7) (Yeates and Clubb, 2007; Kang and Baker, 2011). This provides enhanced stability for Spy0128, which may relate to its biological function as a major pilin subunit.

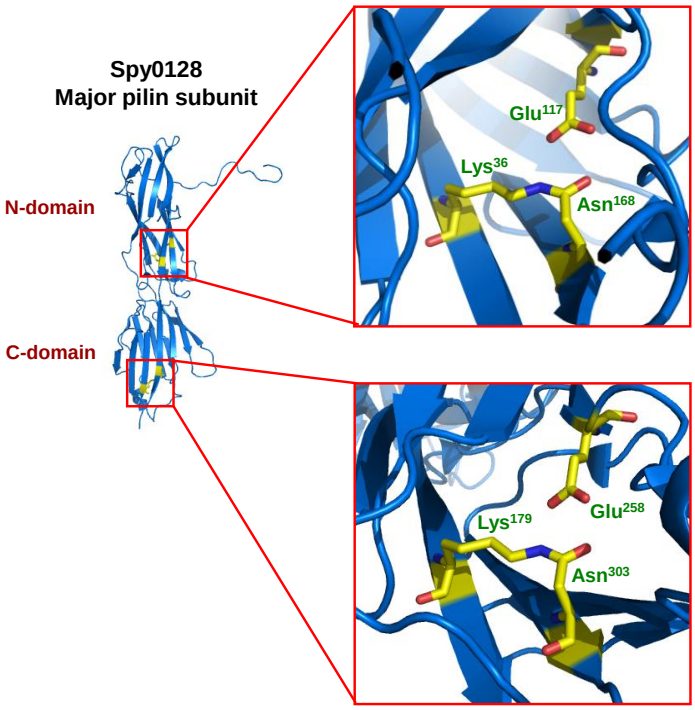


Figure 1.6. Spy0128 contains two spontaneously formed intramolecular isopeptide bonds. Spy0128 (PDB 3B2M) contains two domains composed of all β-sheets, where the first and last β-sheet in each domain is linked by a Lys-Asn isopeptide bond. The N-domain isopeptide bond is formed between Lys^{36} and Asn^{168} and is catalyzed by Glu^{117}, and the C-domain isopeptide bond is formed between Lys^{179} and Asn^{303} and is catalyzed by Glu^{258} (highlighted in yellow).

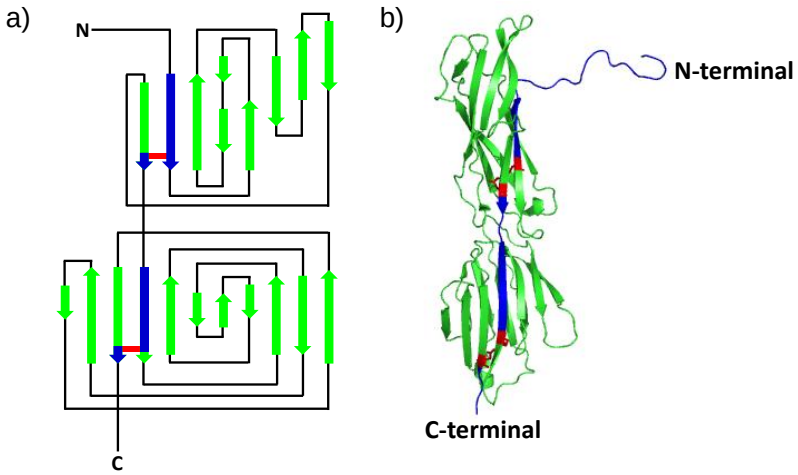


Figure 1.7. Linear covalent connectivity of Spy0128. Isopeptide bonds connect the first and last β-sheet in each domain of Spy0128 creating a linear path of covalent connectivity (blue) in the folded protein structure from the first to the last amino acid. Isopeptide bonds are shown in red. a) A topology diagram of Spy0128 (adapted from (Kang et al., 2007)). b) A structural representation of the covalent connectivity of Spy0128 (PDB 3B2M) (adapted from (Alegre-Cebollada et al., 2010)).

The discovery of intramolecular isopeptide bonds was very surprising, however Kang et al. postulated that these bonds may not be very rare and that they could have simply been mis-annotated in crystal structures in the past (Kang et al., 2007). In crystal structures isopeptide bonds are seen as an area of extended electron density between the side chains of the residues involved and unless one specifically looks for this anomaly, they can be easily overlooked. Therefore, they examined structurally similar proteins to Spy0128 and performed structural similarity searches in the Protein Data Bank using a template of the isopeptide bond forming residues, and they found other examples of proteins containing intramolecular isopeptide that had gone undetected (Kang et al., 2007). All of the proteins were outer-membrane proteins of Gram-positive bacterial pathogens (Kang et al., 2007; Kang and Baker, 2011).

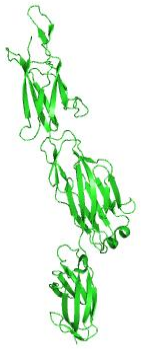
Spontaneously forming isopeptide bonds as a method of covalently linking amino acids is an elegant approach to polypeptide cross-linking. In contrast to the other methods of covalently linking polypeptides that were discussed earlier, spontaneously forming isopeptide bonds have some clear advantages. These include the fact that they do not require enzymatic catalysis or any cofactors, due to the requirement of three reactive residues positioned in the correct conformation surrounded by a hydrophobic environment they are likely to be very specific, an amide bond is chemically stable, and since isopeptide bonds are amide bonds outside of the protein main chain they are likely not going to be subject to proteolytic degradation. Furthermore, isopeptide bond formation is not dependent on the redox status of the environment. Disulfide bond formation within proteins are used by many organisms as a method to enhancing protein stability (Voet and Voet, 2004; Dutton et al., 2008). However, a bioinformatic survey of the genomes of 375 bacteria from diverse backgrounds identified *S. pyogenes* and other bacteria from the same phylum (Firmicutes) as lacking disulfide bonds in exported proteins and also the cellular machinery required to form disulfide bonds (Dutton et al., 2008). Therefore, it is likely that certain Gram-positive bacteria have evolved to incorporate intramolecular isopeptide bonds as a means to stabilize protein structures, and that isopeptide bonds are

likely to have a key physiological role in the function of outer-membrane proteins in these bacteria (Kang et al., 2007; Dutton et al., 2008).

1.3.3. Other intramolecular isopeptide bond containing proteins

In 2007, the crystal structure of Spy0128 was published, providing the first example of a protein containing a spontaneously formed intramolecular isopeptide bond (Kang et al., 2007). This study sparked great interest in the field and many more intramolecular isopeptide bond containing proteins were reported in the years that followed. These newly identified proteins were all outer-membrane proteins from Gram-positive bacteria and included a diverse set of proteins with varying sizes and architectures. A survey of experimentally confirmed intramolecular isopeptide bond containing proteins is provided in Table 1.1, which summarizes key information regarding each protein. A general feature of such proteins is that intramolecular isopeptide bonds are only found in CnaA and CnaB type domains, which are both composed entirely of β-strands (Kang and Baker, 2011). Interestingly, in CnaA type domains the first and penultimate β-strands are linked via an isopeptide bond, whereas in CnaB type domains the first and last β-strands are linked (Kang and Baker, 2011). Furthermore, intramolecular isopeptide bonds can form both between Lys-Asn and Lys-Asp pairs, and the catalytic residue is generally an Asp in CnaA type domains and Glu in CnaB type domains (Table 1.1) (Kang and Baker, 2011). Although not experimentally confirmed, other examples of intramolecular isopeptide bond containing proteins are predicted, such as the minor pilin protein GBS52 *Streptococcus agalactiae* (Krishnan et al., 2007) and the Cna adhesin from *Staphylococcus aureus* (Deivanayagam et al., 2000; Zong et al., 2005), which contained intramolecular isopeptide bonds that were originally unrecognized, but later identified by Kang et al. (Kang et al., 2007). As depicted in Table 1.1, intramolecular isopeptide bonds appear to be a general structural feature of the outer membrane proteins of certain Gram-positive bacteria, and the list of known isopeptide bond

containing proteins will likely increase as more investigations are conducted on this group of proteins.

Protein (PDB Code)	Species	Protein Function	No. of Isopepti-de Bonds	Residues Involved	Size of Prote-in	Additional Comments	Reference
Spy0128 (3B2M) 	*Streptococcus pyogenes*	Major pilin subunit	2	K^{36}-N^{168}-E^{117} K^{179}-N^{303}-E^{258}	294 amino acids	First discovered example of intramolecular isopeptide bonds	(Kang et al., 2007)
BcpA (3KPT) 	*Bacillus cereus*	Major pilin subunit	3	K^{174}-N^{265}-E^{223} K^{273}-N^{383}-D^{312} K^{417}-N^{512}-E^{472}	350 amino acids	4^{th} isopeptide bonds forms in an extended larger structure following pilus assembly	(Budzik et al., 2009)
SpaA (3HR6)	*Corynebacteri-um diphtheriae*	Major pilin subunit	2	K^{199}-N^{321}-D^{241} K^{363}-N^{482}-E^{446}	431 amino acids	Contains a metal binding site and a disulfide bond	(Kang et al., 2009)
RrgA (2WW8) 	*Streptococcus pneumoniae*	Adhesin	2	K^{191}-N^{695}-D^{600} K^{742}-N^{854}-E^{803}	815 amino acids		(Izore et al., 2010)
RrgC	*Streptococcus pneumoniae*	Cell wall anchoring pilin	2	K^{155}-N^{252}-E^{222} K^{264}-N^{354}-E^{322}	347 amino acids	There was no crystal structure provided; isopeptide bonds were confirmed using mass spectrometry	(El et al., 2010)

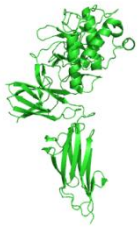
Protein (PDB Code)	Specie	Protein Function	No. of Isopepti-de Bonds	Residues Involved	Size of Prote-in	Additional Comments	Reference
SspB (2WZA)	*Streptococcus gordonii*	Adhesin	2	K^{1082}-N^{1232}-D^{1132} K^{1259}-N^{1393}-D^{1307}	338 amino acids		(Forsgren et al., 2010)
Spy0125 (2XIC) 	*Streptococcus pyogenes*	Adhesin	2	K^{297}-D^{595}-E^{347} K^{610}-N^{715}-E^{680}	430 amino acids	Contains an internal thioester bond between Cys and Gln side chains	(Pointon et al., 2010)
CnaB2 (2X5P)* 	*Streptococcus pyogenes*	A domain of a larger fibronectin binding protein (FbaB)	1	K^{31}-D^{117}-E^{77}	98 amino acids		(Hagan et al., 2010; Oke et al., 2010) and 2X5P manuscri-pt to be published
RrgB (2X9W)	*Streptococcus pneumoniae*	Major pilin subunit	3	K^{193}-N^{318}-D^{241} K^{349}-N^{428}-E^{405} K^{453}-N^{623}-E^{577}	442 amino acids		(Spraggon et al., 2010)
SpaP (3OPU)	*Streptococcus mutans*	Adhesin	2	K^{1157}-N^{1307}-D^{1207} K^{1334}-N^{1469}-D^{1383}	338 amino acids		(Nylander et al., 2011)
GBS80 (3PF2)	*Streptococcus agalactiae*	Major pilin subunit	2	K^{210}-N^{351}-D^{250} K^{387}-N^{515}-E^{471}	317 amino acids		(Vengadesan et al., 2011)
AgI/II (3QE5)	*Streptococcus mutans*	Adhesin	3	K^{1006}-N^{1121}-D^{1050} K^{1161}-N^{1311}-D^{1211} K^{1338}-N^{1473}-D^{1387}	487 amino acids		(Larson et al., 2011)
PitB	*Streptococcus pneumoniae*	Major pilin subunit	2	K^{63}-N^{214}-E^{164} K^{243}-N^{372}-E^{325}	410 amino acids (Acce-ssion #: ADD91 687)	There was no crystal structure provided; isopeptide bonds were confirmed using mass spectrometry	(Zahner et al., 2011)
FimA (3QDH)	*Actinomyces oris*	Major pilin subunit	2	K^{206}-N^{319}-D^{246} K^{363}-N^{485}-E^{450}	290 amino acids		(Mishra et al., 2011)

Table 1.1. A survey of experimentally confirmed intramolecular isopeptide bond containing proteins. Intramolecular isopeptide bonds are a general feature of outer-membrane proteins of certain Gram-positive bacteria. This table demonstrates the diversity in shape and size of the proteins, and the amino acids used to form isopeptide linkages. The size of proteins shown is the length based on the available crystal structures in the Protein Data Bank (PDB). The structures of certain proteins are shown to illustrate structural diversity. *Note that although the CnaB2 protein has been published as PDB 2X5P, we obtained this construct and its crystal structure from the authors prior to publication of the crystal structure.

1.3.4. Mechanism of autocatalytic isopeptide bond formation

Intramolecular isopeptide bonds likely form as the nascent polypeptide chains fold following protein expression (Hu et al., 2010). As such, reaction intermediates cannot be easily isolated to allow for mechanistic studies of isopeptide bond formation at the molecular level. Accordingly, based on available crystal structures and mutagenesis studies a reaction mechanism for autocatalytic isopeptide bond formation has been proposed (Helgstrand et al., 2003; Kang et al., 2007), and has subsequently been supported by quantum mechanical/molecular mechanical studies (Kang and Baker, 2011; Hu et al., 2010; Hagan et al., 2010).

Autocatalytic isopeptide bond formation between Lys and Asn/Asp side chains requires a nearby general base (Glu/Asp) to catalyze the reaction by mediating the transfer of protons (Kang and Baker, 2011). The requirement of the general base for catalysis has been experimentally confirmed, where mutation of the catalytic Glu residues in both the N- and C-terminal domains of Spy0128 (Fig. 1.6) to Ala led to the loss of isopeptide bond formation (Kang et al., 2007). Similar results have been found in other intramolecular isopeptide bond forming proteins including BcpA (Budzik et al., 2009), RrgA (Izore et al., 2010), and CnaB2 (Hagan et al., 2010). Furthermore, isopeptide bond formation is proximity-induced and hence the three residues involved in the reaction must have the proper geometry (Kang and Baker, 2011; Hu et al., 2010). Finally, the reacting residues must be in a hydrophobic environment which changes the pKa values of the amino acid side chains resulting in an inverse protonated state, where the Lys side chain is unprotonated and the Glu/Asp side chain is protonated (Kang and Baker, 2011; Hu et al., 2010).

The mechanism of autocatalytic isopeptide bond formation proceeds in several stages (Fig. 1.8): (i) the Cγ carbonyl group of the Asn/Asp is polarized via hydrogen bonding to the protonated catalytic Glu/Asp residue (Kang et al., 2007; Kang and Baker, 2011), (ii) the unprotonated Lys ε-amino group nucleophilically attacks the polarized Cγ carbonyl group of Asn/Asp resulting in the formation of a tetrahedral intermediate (Hagan

et al., 2010; Kang et al., 2007; Kang and Baker, 2011; Hu et al., 2010), (iii) a proton is transferred from the Lys ε-amino group to the catalytic Glu/Asp residue (Kang and Baker, 2011), and (iv) a proton is transferred from the catalytic Glu/Asp residue to the leaving group of the Asn/Asp (Kang and Baker, 2011). Subsequently, the formation of a Lys-Asn isopeptide bond results in the release of a NH_3 molecule, while a Lys-Asp linkage leads to the release of a H_2O molecule (Fig. 1.8) (Kang and Baker, 2011).

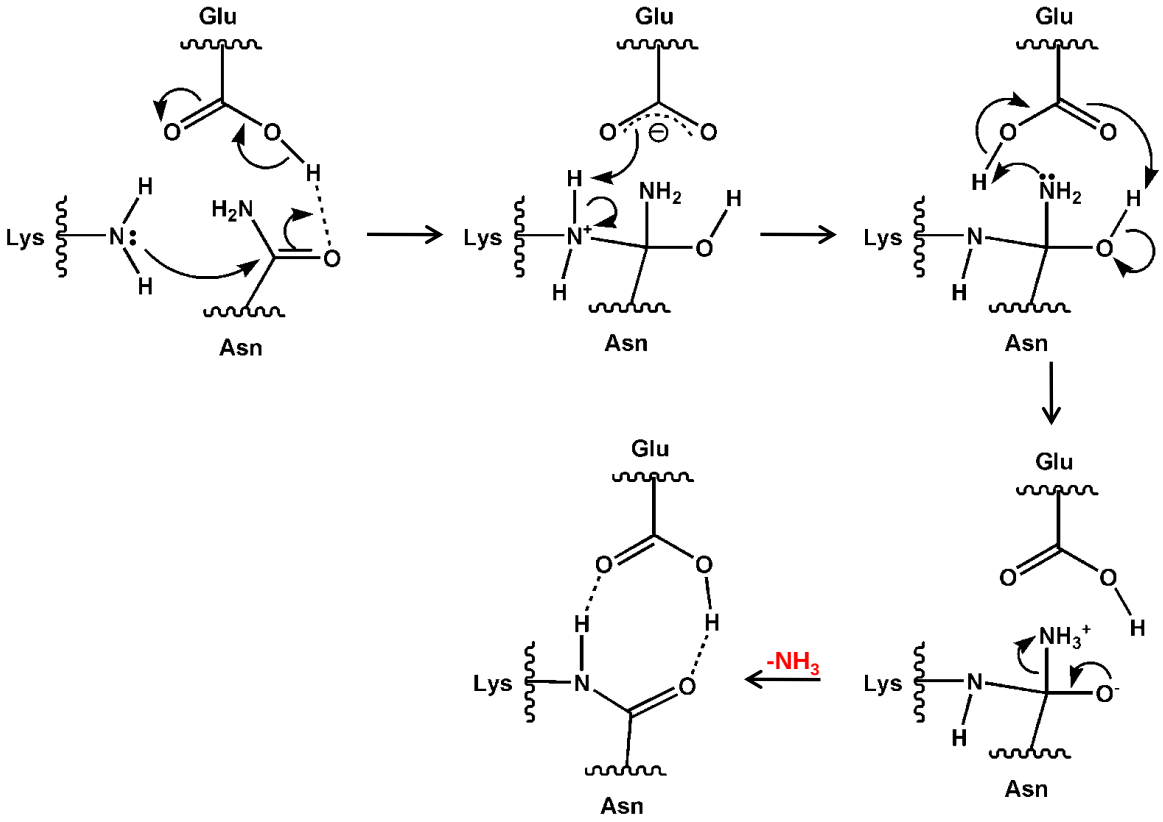


Figure 1.8. Mechanism of autocatalytic isopeptide bond formation. Isopeptide bond formation between a Lys and an Asn side chain is illustrated, with a nearby Glu residue to catalyze the reaction (based on the mechanism proposed by Kang and Baker (Kang and Baker, 2011)). Hydrogen bonding (dashed lines) by the protonated Glu side chain polarizes the Cγ carbonyl of Asn. The unprotonated Lys ε-amino group nucleophilically attacks the Cγ carbonyl of Asn resulting in the formation of a tetrahedral intermediate. The catalytic Glu shuttles a proton from the Lys ε-amino group to the NH_2 leaving group of Asn. Subsequently, the Asn releases NH_3 resulting in the formation of an isopeptide bond, which is also hydrogen bonded to the catalytic Glu. An Asp residue can also catalyze this reaction and isopeptide bond formation can similarly occur between a Lys and an Asp residue resulting in the release of a H_2O molecule. This figure was adapted from (Kang and Baker, 2011).

1.4. Gram-positive bacterial pathogens use intramolecular isopeptide bonds to stabilize outer membrane proteins

1.4.1. Pilus assembly in Gram-positive bacteria

Gram-positive bacterial pathogens are responsible for many human ailments from mild conditions such as skin infections and pharyngitis, to severe diseases such as necrotizing fasciitis and pneumonia (Molinari and Chhatwal, 1999). Critical to the infective process is the ability of the bacteria to recognize and bind to the correct host tissues, and to aid in this process the bacteria have evolved complex outer-membrane structures (Telford et al., 2006). One particular class of outer-membrane structures that are important for adhering to host cells are special hair-like appendages called pili, which are long and thin attachments that extend from the surface of bacteria and have receptor-specific adhesins at their tips to mediate binding to cellular targets (Telford et al., 2006). Pili are important to many events in the infective process from initial recognition and adhesion to host tissues, colonization of host tissues through the recruitment of other bacteria, and eventual biofilm formation to aid in evasion of host defences (Kline et al., 2010; Mandlik et al., 2008; Telford et al., 2006). In mediating bacterial-host interactions, pili are exposed to harsh extra-cellular conditions where fluctuation in temperature and pH can occur (Kang and Baker, 2011; Kang and Baker, 2009; Hagan et al., 2010). Furthermore, as pili adhere to host surfaces they are exposed to large amounts of tensile forces (Kang and Baker, 2011; Alegre-Cebollada et al., 2010). Therefore, pili must be assembled in a way to withstand such harsh conditions and made from proteins that are stable enough to maintain their structural integrity under extreme conditions.

To impart extreme structural stability to their pili, Gram-positive bacteria utilize covalent cross-linking via isopeptide bonds during pilus assembly (Alegre-Cebollada et al., 2010; Kang and Baker, 2011). Pili are assembled through the covalent polymerization of a single major pilin subunit by sortases, a versatile class of transpeptidase enzymes that can utilize a diverse array of protein substrates (Telford et al., 2006). As a result, the pilus is one protein molecule thick but can grow up to 3 μm in length (Telford et al., 2006).

Subsequently, an accessory factor such as an adhesin is attached covalently to the tip of the polymerized pilus to mediate binding to a cellular receptor (Kline et al., 2010). The first step in pilus assembly is that the pilin subunits are expressed containing a signal sequence for Sec-dependent translocation across the membrane, a membrane spanning domain to allow the subunit to anchor in the membrane, and an LPXTG motif (Telford et al., 2006; Kline et al., 2010). Once the pilin subunits are anchored in the membrane, the next step in pilus assembly is the recognition and cleavage of the LPXTG motif by membrane anchored sortases (Kline et al., 2010). Different sortases are required to polymerize the pilin subunits and to covalently anchor the final pilus structure to the cell wall (Kline et al., 2010). During the assembly process, a sortase cleaves the LPXTG motif between the threonine-glycine bond with a reactive thiol group in its active site, thus forming a thioester bond between a cysteine residue in the enzyme and the carboxyl group of the exposed threonine (Fig. 1.9) (Telford et al., 2006; Kline et al., 2010). This creates an acyl-enzyme intermediate that is involved in the polymerization of the pilus, where nucleophlic attack of the thioester bond by an ε-amino group from the side chain of a specific lysine residues leads to extension of the pilus (Fig. 1.9) (Telford et al., 2006). This process is repeated depending on the abundance of pilin subunits until pilus polymerization ceases (Telford et al., 2006). In the final step, the polymerized pilus is covalently attached to the cell wall in a reaction that is catalyzed by a housekeeping sortase (Kline et al., 2010) (Fig. 1.9). This reaction proceeds through nucleophilic attack of the acyl-enzyme intermediate by an amino group of the pentaglycine cross-bridge of the lipid II precursor in the peptidoglycan layer (Fig. 1.9) (Telford et al., 2006; Mandlik et al., 2008; Hendrickx et al., 2011). The attachment of the accessory factors which mediate binding of the pilus to cellular receptors is not well understood but is believed to proceed through a similar sortase dependent mechanism due to the presence of LPXTG motifs (Telford et al., 2006). Therefore, the final pilus structure can be thought of as a single linear polypeptide chain that is connected covalently from the cell wall to the pilus tip.

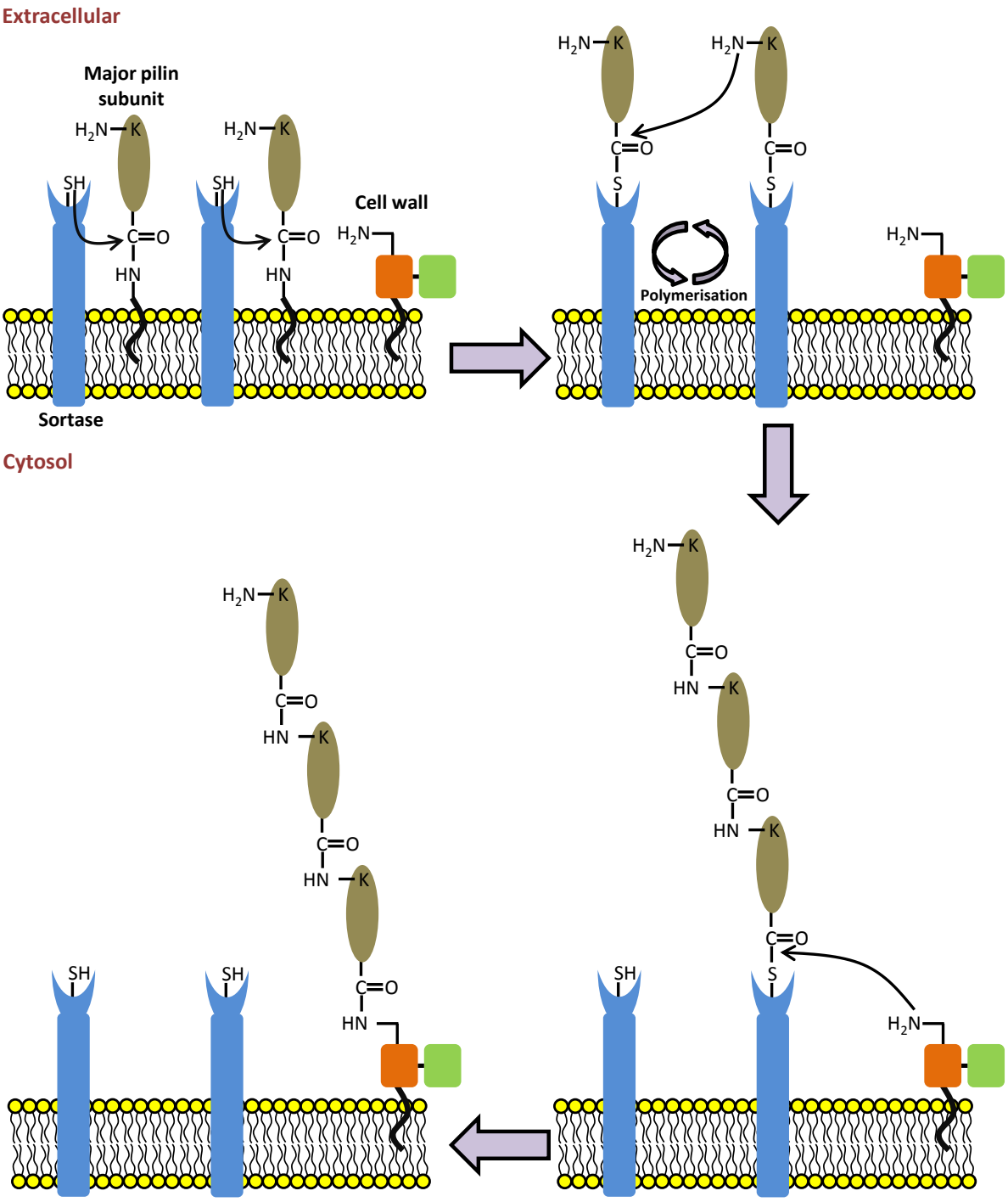


Figure 1.9. A schematic representation of sortase mediated assembly of Gram-positive bacterial pili. The major pilin subunit is translocated across the membrane through a Sec-dependent mechanism and is then anchored in the membrane. A cysteine thiol group from a membrane-anchored sortase cleaves the peptide bond between threonine and glycine in the LPXTG motif of the pilin, forming an acyl-enzyme intermediate between the thiol group and the carboxyl group of threonine. Subsequent nucleophilic attack of the acyl-enzyme intermediate by an ε-amino group from a conserved lysine residue extends the pilus by one subunit via an isopeptide bond. This polymerization reaction repeats many times depending on the abundance of the pilin subunits anchored in the membrane. Once pilus polymerization is complete, nucleophilic attack by an amino group from the pentaglycine cross-bridge of the lipid II precursor on the acyl-enzyme intermediate transfers the pilus to the peptidoglycan layer via an

isopeptide bond. The cell wall components N-acetyl glucosamine is shown as an orange box and N-acetyl muramic acid is shown as a green box. This figure was adapted from (Telford et al., 2006).

1.4.2. Biological function of intramolecular isopeptide bonds

Pili are important virulence factors that are used by pathogenic bacteria to bind and colonize host tissues, which makes them desirable targets for therapeutic intervention and vaccine development (Mora et al., 2005). Interestingly Gram-negative and Gram-positive bacteria have evolved to produce their pili through different mechanisms even though they serve the same functions. Gram-negative bacterial pili are produced through non-covalent linking of pilin subunits, which are structurally stabilized through disulfide bonding and the bundling of multiple subunits to from thicker pili (Kline et al., 2010; Yeates and Clubb, 2007; Craig et al., 2004). In contrast, Gram-positive bacterial pili utilize intramolecular isopeptide bonds (Fig. 1.7) to stabilize pilin subunits which are then cross-linked and tethered to the cell wall via intermolecular isopeptide bonding (Fig. 1.9) (Kline et al., 2010; Kang and Baker, 2011).

Although Spy0128 (*S. pyogenes* major pilin subunit) was only recently demonstrated as being resistant to trypsin digestion due to the presence of intramolecular isopeptide bonds, it was first identified over half a century ago as a result of its resistance to trypsin digestion where it was classified as the Lancefield-T (trypsin-resistant) antigen and used for serotyping pathogenic *Streptococci* (Mora et al., 2005; Lancefield and Dole, 1946; Kang et al., 2007; Kang and Baker, 2009). Intramolecular isopeptide bonding also enhances the thermodynamic stability of the protein by increasing the protein melting temperature by 30°C (Kang and Baker, 2009). Further investigations on Spy0128 using single molecule force spectroscopy have demonstrated that due to the presence of the intramolecular isopeptide bonds, the protein was mechanically inextensible making it the most mechanically stable protein ever studied (Alegre-Cebollada et al., 2010).

The highly stable Spy0128 major pilin subunit and its sortase-mediated covalent cross-linking during pilus assembly paints an elegant picture of the structure and function

of *S. pyogenes* pili. The individual Spy0128 subunits are elongated protein structures composed of two domains containing only β-sheets (Kang et al., 2007). In each domain the first and last β-sheet is connected by an isopeptide bond between a Lys and an Asn side chain, and these isopeptide bonds are strategically positioned at the end of each domain (Kang et al., 2007). This creates a linear path of covalent connectivity through the protein (Fig. 1.7) (Kang et al., 2007; Yeates and Clubb, 2007). Subsequently, during sortase assembly of the pilus each pilin subunit is covalently linked to another through an isopeptide bond between a cleaved LPXTG motif and a conversed Lys (Fig. 1.9) (Telford et al., 2006). This conserved Lys in Spy0128 is situated near the top of the N-domain and when cross-linked to another subunit it forms a head-to-tail packing of the pilins (Kang et al., 2007). Therefore, when the complete pilus structure is formed a single linear path of covalent connectivity can be traced from the cell wall anchored first subunit to the final subunit at the pilus tip (Yeates and Clubb, 2007). This provides covalent polymerization of a protein that is inextensible to create a pilus that can withstand the large amounts of mechanical force that it experiences during binding events (Kang and Baker, 2011; Alegre-Cebollada et al., 2010). However, it is of note that the N-domain isopeptide bond is not required for this linear covalent path through the polymerized pilin since pilin subunits are linked through a conserved Lys residue. This residue is situated at the base of the last β-strand in the N-terminal domain, therefore negating the need for linking the first and last β-strands in the domain (Alegre-Cebollada et al., 2010).

This notion of covalent connectivity was extended even further by the Banfield group who discovered that Spy0125, the pilus adhesin that is covalently attached to Spy0128 and is displayed at the tip of the pilus to mediate host binding, contains an internal thioester bond, which forms between the side chains of a Cys and a Gln residue inside a surface exposed shallow cleft (Pointon et al., 2010). Mutation of the Cys residue involved in the thioester bond revealed that the thioester bond did not contribute to protein structural stability, however the mutant was compromised in binding to host cells (Pointon et al., 2010). This led the authors to propose a model of pilus-mediated bacterial

interaction with the host, where the exposed thioester moiety can react with a nucleophile on the host to form a covalent linkage between the adhesin and the host (Pointon et al., 2010). The adhesin is in turn covalently linked to Spy0128, which forms a covalently linked pilus shaft and the pilus shaft is also covalently linked to the bacterial cell wall (Pointon et al., 2010). Therefore, the complete bacterium-host interaction is mediated through covalent bonds.

Although intramolecular isopeptide bonds appear to be commonly used by Gram-positive bacteria as stabilizing factors, we have only recently begun to identify and understand these bonds. Further investigations of other outer-membrane proteins of Gram-positive bacteria will shed light on this amazing feature of amino acid chemistry, and provide insight in to how we could exploit this reaction to develop new technologies.

1.5. Protein splitting technology

Autocatalytic isopeptide bond formation represents a novel method of linking polypeptide sequences and may be adapted for the development of covalently interacting peptide-protein pairs. However, a critical factor in this approach would be reproducing the specific environment required for this reaction to occur, namely a hydrophobic surrounding with a specifically positioned Glu/Asp to mediate the reaction (Kang and Baker, 2011). In isopeptide bond containing proteins this environment is produced intramolecularly and requires the correct folding of the protein (Kang and Baker, 2011; Hu et al., 2010). Therefore, to construct our desired peptide-protein pair we must develop a system that reproduces the intramolecular environment of these proteins using intermolecularly interacting peptides and proteins.

Split protein technology is a technique where a protein can be split into two polypeptide fragments. The two polypeptide fragments can then interact in solution and spontaneously refold in to the original protein structure and regain protein function (Ghosh et al., 2000; Ozawa, 2006). This approach has been well established in the literature and demonstrated on various protein scaffolds including RNase (Richards, 1958), aminoacyl-

tRNA synthetase (Shiba and Schimmel, 1992), and chymotrypsin inhibitor-2 (de Prat and Fersht, 1994) among others (Ghosh et al., 2000). Furthermore, this technology has also been applied for the development of reporter systems for the study of protein-protein interactions (Ozawa, 2006). In 1994, Johnsson and Varshavsky were the first group to use this approach to develop a sensor for protein-protein interactions, which they called ubiquitin-based split-protein sensor (USPS) (Johnsson and Varshavsky, 1994; Ozawa, 2006). They demonstrated the utility of USPS for the *in vivo* study of protein-protein interactions with temporal resolution (Johnsson and Varshavsky, 1994). To develop USPS, they split ubiquitin in to two fragments and mutated the N-terminal fragment so that efficient reconstitution could not be achieved with the C-terminal fragment (Johnsson and Varshavsky, 1994). They further fused a sensor protein to the C-terminal fragment to provide a means of detecting protein-protein interactions (Johnsson and Varshavsky, 1994). Individually the two split domains would not reconstitute in solution, but when they were linked to proteins that interacted with each other, the two ubiquitin fragments would be positioned in close proximity which would induce reconstitution (Johnsson and Varshavsky, 1994). The reconstituted ubiquitin could then be recognized by an ubiquitin-specific protease that would cleave the sensor protein (fused to the C-terminal ubiquitin fragment), resulting in a detectable signal for the protein-protein interaction (Johnsson and Varshavsky, 1994). Similar approaches were later used to develop reporter systems based on other split proteins including inteins (Xu and Evans, Jr., 2005), Green Fluorescent Protein (Ghosh et al., 2000), dihydrofolate reductase (Singh et al., 2006), and others (Ozawa, 2006).

Protein splitting is illustrated in Figure 1.10 with a hypothetical fluorescent protein, where a protein can be genetically split in to two separate fragments that can be expressed and purified as separate protein fragments *in vitro* or simultaneously expressed *in vivo*. When the two unstructured polypeptides are present together in solution, they spontaneously interact and fold in to the correct protein structure and regain protein function (Fig. 1.10). The effectiveness of this approach is dependent on two main

concerns which are the stability and affinity of the split constructs. When a protein is split and is in a disordered state, it is more vulnerable to protease degradation which can reduce its lifetime. In addition, there may be difficulties in recombinant expression of split proteins because when a protein is fragmented and in a disordered state, its hydrophobic regions are exposed to an aqueous environment which can lead to protein aggregation and precipitation. The affinity of the two fragments is also a major concern because they must be able to recognize and bind each other in solution. This is dependent on the conformation of the unstructured polypeptides in solution, and requires them to be in a receptive state for folding. If they form hairpin structures or aggregates then it is unlikely for them to be able to reconstitute.

In using this approach to develop a peptide-protein interacting pair, the issue of affinity is even further heightened. As discussed in Section 1.1, the affinity between a peptide and a protein binding partner is limited due to the structural flexibility of the peptide and its small surface area available for binding (Houk et al., 2003). Therefore, in splitting a protein to develop a peptide that can recognize and bind the rest of the protein fragment, a major consideration would be the length of the peptide required to mediate high affinity and specific binding.

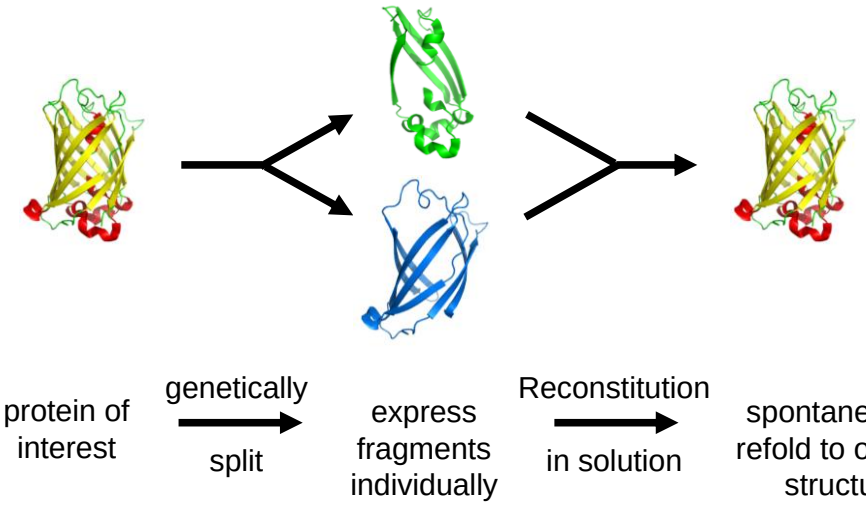


Fig. 1.10. Splitting of a hypothetical fluorescent protein. A hypothetical fluorescent protein is genetically split in to two separate sequences, which can be individually expressed and purified. When the two protein fragments are mixed in solution they come

together and spontaneously fold in to the original protein structure and regain protein function. This figure was adapted from (Ozawa, 2006).

1.6. Objectives of this work

Herein, I was intending to develop a new class of peptide fusion tags that could be genetically encoded and were able to bind to their protein partners through an irreversible covalent bond without the requirement of any accessory factors. To develop these peptide fusion tags, I aimed to take advantage of the spontaneously formed intramolecular isopeptide bonds present in certain outer-membrane proteins of Gram-positive bacteria. By splitting a protein strategically to form a small peptide fragment and a larger protein binding partner each containing one residue involved in isopeptide bond formation, I hypothesized that the two polypeptides would be able to recognize and interact with each other in solution. The peptide-protein pair would then fold in to the natural protein structure, which would facilitate the formation of an intermolecular isopeptide linkage between the peptide and protein fragments in a reaction catalyzed by a Glu/Asp residue present in the protein partner.

Proof of principle experiments were conducted on the well studied Spy0128 major pilin protein from *S. pyogenes*. This led to the development of the isopeptag:pilin-C peptide-protein pair that was able to covalently reconstitute *in vitro* under most biological relevant conditions with a maximum yield of ~60% after 72 hr. The isopeptag:pilin-C pair was also able to specifically reconstitute *in vivo* inside the complex environment of a bacterial cytosol and be used for cellular imaging applications. A more efficient covalently reconstituting peptide-protein pair was developed by splitting the CnaB2 protein, where almost complete covalent complex formation was achieved after 24 hr. Rational protein modifications were then performed to optimize the split CnaB2 peptide-protein pair resulting in the development of the SpyTag:SpyCatcher pair. Characterization of the SpyTag and SpyCatcher pair revealed a peptide and protein partner that could efficiently react, where covalent complex formation proceeded to almost completion within 2 hr.

Furthermore, the SpyTag:SpyCatcher pair was able to efficiently react under all experimental conditions tested and could specifically reconstitute on the surface of mammalian cells.

Chapter 2:

Methods

Parts of this chapter have been published in (Zakeri and Howarth, 2010), which is provided in the appendix.

Chapter 2: Methods

2.1. Cloning

2.1.1. General cloning procedures

All PCR experiments were performed using KOD Hot Start DNA polymerase (Novagen) based on manufacturer recommended conditions. Briefly, 50 µL PCR reactions were prepared in MilliQ H_2O containing KOD Hot Start DNA polymerase buffer, 1.5 mM $MgSO_4$, 0.2 mM dNTPs, 2 µM sense primer, 2 µM anti-sense primer, ~1 ng/µL template DNA, and 0.2 U/ µL KOD Hot Start DNA polymerase. PCR reactions were then placed in a Bio-Rad C1000 thermal cycler with the following cycling conditions: 3 min at 95°C, then 30 s at 95°C, 1 min at 55°C, 1 min at 72°C for 25 cycles, then 2 min at 72°C, followed by incubation at 10°C. All PCR reactions were optimized by varying the annealing temperatures used. DNA fragments were all analyzed on 0.7–2% agarose gels (depending on the size of the DNA fragment) run in an Owl electrophoresis system (Thermo Scientific) for 40 min at 130 V. For ligation experiments, PCR products and vectors were digested with the appropriate restriction enzymes under manufacturer recommended conditions and then agarose gel purified using a QIAquick Gel Extraction Kit (Qiagen). Ligation reactions were prepared by mixing 3:1, 1:1, and 1:3 ratios of digested PCR products and vectors in 10 µL reactions containing T4 DNA ligase buffer and 20 U/µL T4 DNA Ligase (New England Biolabs), and the reactions were then incubated at 25°C for 1 hr. Subsequently, 4 µL of each ligation reaction was transformed in to 50 µL of heat-competent *Escherichia coli* DH5α (Invitrogen) or XL-1 blue (Stratagene) cells.

Site-directed mutagenesis experiments were performed using the QuikChange™ (Stratagene) protocol. Briefly, 25 µL PCR reactions were prepared in MilliQ H_2O containing KOD Hot Start DNA polymerase buffer, 1.5 mM $MgSO_4$, 0.4 mM dNTPs, 0.4 µM sense primer, 0.4 µM anti-sense primer, 5% dimethyl sulfoxide (DMSO), ~0.2 ng/µL

template DNA, and 0.2 U/ µL KOD Hot Start DNA polymerase. The cycling conditions used for site-directed mutagenesis experiments were: 3 min at 94°C, then 20 s at 94°C, 1 min at 55°C, 4 min at 68°C for 20 cycles, followed by incubation at 10°C. PCR reactions were optimized by varying the annealing temperatures. PCR reactions (5 µL) were analyzed using agarose gels and if a PCR product was visible at the appropriate weight, then 1 U/µL DpnI (New England Biolabs) was added to the PCR products (20 µL) and the reactions were incubated at 37°C for 4 hr. Subsequently, 4 µL of each ligation reaction was transformed in to heat-competent *E. coli* as described above.

All constructs made in this study were verified by DNA sequencing (Source BioScience LifeSciences).

2.1.2. Cloning of Spy0128 constructs

The numbering of residues in all of the constructs made herein are based on the numbering of the Spy0128 protein (PDB 3B2M) (Kang et al., 2007). To clone the *Spy0128* gene, the primers 5'–GGGGCATATGGCTACAACAGTTCACGGG–3' and 5'–GGGGAA GCTTTTATTCAAAGTCTTTTTTATTTG–3' were used to PCR amplify the gene from *Streptococcus pyogenes* M1 GAS strain SF370 genomic DNA (ATCC 700294D-5). The amplified *Spy0128* gene was then inserted in to the expression vector pET28a (Novagen) using *NdeI* and *HindIII* restriction sites. To generate Spy0128 E117A, the primer 5'–GTGTTTATTATTACAAAGTAACTGCGGAGAAGATAGATAAAGTTCCTGG–3' (mutated residue is underlined here and thereafter) and its reverse complement were used to mutate Glu^{117} to Ala in pET28a-Spy0128. To create Spy0128 K179A, the primer 5'–TCT ACTACATTAACGGTGAAGGCAAAAGTTTCAGGTACCGGTGG–3' and its reverse complement were used to mutate Lys^{179} to Ala in pET28a-Spy0128. The isopeptide null Spy0128 E117A K179A mutant was generated by introducing a K179A mutation in to pET28a-Spy0128 E117A using the primer 5'–TCTACTACATTAACGGTGAAGGCAAAAG TTTCAGGTACCGGTGG–3' and its reverse complement.

MBP-isopeptag-N was generated by amplifying the maltose binding protein (MBP) from the pMAL vector (New England Biolabs) using the primers 5'–CAAGCATATGAAAAT CGAAGAAG–3' and 5'–CGAACCCGAGCTCGAATTAGTCTG–3', while the peptide isopeptag-N (Spy0128 residues 18-45) was amplified using the primers 5'– CTAATTCGA GCTCGGGTTCGGGTGAAAGTGGTGCTACAACAGTTCACGGG–3' and 5'–GGGGAAG CTTTTATGCATTGCTATTAACTAAATC–3' from pET28a-Spy0128. Subsequently, overlap-extension PCR was used to link the two fragments, which were then inserted in to the expression vector pET28a using *NdeI* and *HindIII* restriction sites. To construct pilin-N (Spy0128 residues 46-308), the primers 5'–GGGGCATATGGGATTAATTCCAAATACAG AT–3' and 5'–GGGGAAGCTTCTAGTGATGGTGATGGTGATGTCCTGATCCTTCAAAGT CTTTTTTATTTG–3' were used to amplify the *pilin-N* sequence from pET28a-Spy0128, which was then inserted in to the expression vector pET28a using *NdeI* and *HindIII* restriction sites. To generate pilin-N E117A, the primer 5'–GTGTTTATTATTACAAAG TAACTGCGGAGAAGATAGATAAAGTTCCTGG–3' and its reverse complement were used to mutate Glu^{117} to Ala in pET28a-pilin-N.

CB-MBP was constructed using the primers 5'–GGGGCATATGGGAACTGATAAA GATATGACC–3' and 5'–ACCACTTTCACCACTACCTTCAAAGTCTTTTTTATTTG–3' to amplify CB (Spy0128 residues 293-308) from pET28a-Spy0128, while the primers 5'–AA GGTAGTGGTGAAAGTGGTAAAATCGAAGAAGGTAAA–3' and 5'–GGGGAAGCTTTTAC GAGCTCGAATTAGTCTG–3' were used to amplify MBP from the pMAL vector. Subsequently, the two fragments were linked via overlap-extension PCR and inserted in to the pET28a vector using *NdeI* and *HindIII* restriction sites. This resulted in the positioning of a 6 residue linker (GSGESG) between CB and MBP. To construct isopeptag-MBP, a F307A mutation was introduced in to pET28a-CB-MBP using the primer 5'–TTTACAAATAAAAAAGACGCTGAAGGTAGTGGTGAAAG–3' and its reverse complement. The 6X His tag sequence from pET28a-CB-MBP was removed using the QuikChange™ (Stratagene) protocol and the primer 5'–TAAGAAGGAGATATACCATGG GCACTGATAAAGATATGACCATTAC–3' and its reverse complement to generate

CBΔHis-MBP. Isopeptag-MBP -5 (Spy0128 residues 298-308) was constructed using the primers 5'–GGGGCATATGGGAACCATTACTTTTACAAAT–3' and 5'–GGGGAAGCTTTTACGAGCTCGAATTAGTCTG–3' to amplify the gene from pET28a-isopeptag-MBP, which was then inserted in to pET28a using *NdeI* and *HindIII* restriction sites. Pilin-C (Spy0128 residues 18-299) was made using the primer 5'–ACAAGAGACATCTACTGATAAAGATATGACCATTTAGTTTACAAATAAAAAAGACTTTGAATAAAAGCTTG–3' and its reverse complement to change residue 300 in pET28a-Spy0128 to a stop codon. To construct pilin-C K179A where the isopeptide forming Lys^{179} was changed to Ala, the primer 5'–ACAAGAGACATCTACTGATAAAGATATGACCATTTAGTTTACAAATAAAAAAGACTTTGAATAAAAGCTTG–3' and its reverse complement were used to introduce a stop codon at residue 300 in pET28a-Spy0128 K179A. Pilin-C -7 (Spy0128 residues 18-292) was made using the primer 5'–GCAGGTAATTCAACTGAACAAGAGACATCTTAGGATAAAGATATGACCATTACTTTTACAAAT–3' and its reverse complement to introduce a stop codon at residue 293 in pET28a-Spy0128. Pilin-C -2 (Spy0128 residues 18-297) was constructed using the primer 5'–ACTGAACAAGAGACATCTACTGATAAAGATATGTAGATTACTTTTACAAATAAAAAAGACTTTGAAGTG–3' and its reverse complement to change residue 298 to a stop codon in pET28a-Spy0128. Pilin-C -1 (Spy0128 residues 18-298) was made using the primer 5'–GAACAAGAGACATCTACTGATAAAGATATGACCTAGACTTTTACAAATAAAAAAGACTTTGAATAAAAG–3' and its reverse complement to change residue 299 in pET28a-Spy0128 to a stop codon. Pilin-C +1 (Spy0128 residues 18-300) was constructed using the primer 5'–CAAGAGACATCTACTGATAAAGATATGACCATTACTTAGACAAATAAAAAAGACTTTGAATAAAA–3' and its reverse complement to change residue 301 in pET28a-Spy0128 to a stop codon.

For co-expression of isopeptag-MBP and pilin-C or pilin-C K179A inside *E. coli*, a bicistronic vector was generated using primers 5'–GGGGAAGCTTCTCTAGAAATAATTTTGTTTAAC–3' and 5'–GGGGCTCGAGCTAAATGGTCATATCTTTATC–3' to amplify the *pilin-C* gene from pET28a-pilin-C and the *pilin-C K179A* gene from pET28a-pilin-C K179A. Subsequently, the PCR products were inserted in to pET28a-isopeptag-MBP

using *HindIII* and *XhoI* restriction sites. This resulted in the generation of the pET28a-(isopeptag-MBP + pilin-C) and pET28a-(isopeptag-MBP + pilin-C K179A) expression vectors, where an mRNA would be transcribed from a single T7 promoter that coded for isopeptag-MBP-spacer-Shine-Dalgarno sequence-pilin-C/K179A. For mammalian cell surface display of isopeptag, the construct isopeptag-CFP-TM was made using the primers 5'–GACAGATCTGGCGGCACTGATAAAGATATGACCATTACTTTTACAAATAAA AAAGACTTTGAAGGTAGTGGTATGGTGAGCAAGGGCGAG–3' and 5'–ACTCTCGGCA TGGACGAGCTATACAAGCGGCCGCGGGGAG–3' to amplify CFP from the construct AP-CFP-TM (Chen et al., 2005), which was then inserted in to the mammalian expression vector pDisplay (Invitrogen) using *BglII* and *SacII* restriction sites.

2.1.3. Cloning of ACE19 constructs

The sequence and numbering of residues of all ACE19 constructs were based on the sequence provided with the crystal structure of the protein (PDB 2OKM), and is shown below:

```
34     40         50         60         70         80         90
ERDYPF FYKVGDLAGE SNQVRWFLNV NLNKSDVTED ISIADRQGSG QQLNKESFTF DIVNDKETKY
100        110        120        130        140        150
ISLAEFEQQG YGKIDFVTDN DFNLRFYRDK ARFTSFIVRY TSTITEAGQH QATFENSYDI
160        170        180
NYQLNNQDAT NEKNTSQVKN V
```

To clone wild-type ACE19 (residues 34-180), which is a subdomain of the outer-membrane protein ACE from *Enterococcus faecalis* (Ponnuraj et al., 2002; Ponnuraj and Narayana, 2007) the primers 5'–GGGGCATATGGAGCGAGACTATCCGTTTTT–3' and 5'–GGGGAAGCTTCTAGTGATGGTGATGGTGATGTCCTGATCCCTAAACATTTTTAA CCTGTGATG–3' were used to amplify the *ACE19* gene from *E. faecalis* (Andrewes and Horder 1906) cells (Health Protection Agency Culture Collections, UK). The PCR product was then inserted in to the pET28a expression vector using *NdeI* and *HindIII* restriction sites. To construct ACEA-MBP, the primers 5'–GGGGCATATGGGAGAGCGAGACT

ATCCGTTTTTTTATAAAGTAGGCGATGGTAGTGGTAAAATCGAAGAAGGTAAA–3’ (containing the ACEA sequence with a GSG linker followed by the primer sequence for MBP) and 5’–GGGGAAGCTTTTACGAGCTCGAATTAGTCTG–3’ were used to amplify MBP from the template pET28a-PepB-MBP (described in section 2.1.4). The PCR product was then inserted in to pET28a using *NdeI* and *HindIII* restriction sites. To create ACE44 (ACE19 residues 44-180), the primer 5’– CGCGGCAGCCATATGGGCGATTT GGCTG–3’ and its reverse complement were used to delete the appropriate sequence in pET28a-ACE19. To generate ACE50 (ACE19 residues 50-180), the primer 5’–CGCGCG GCAGCCATATGTCAAATCAAGTACGTTG–3’ and its reverse complement were used to delete the appropriate sequence in pET28a-ACE19.

2.1.4. Cloning of CnaB2 constructs

The sequence and numbering of the residues for all CnaB2 constructs is shown below and is based on the sequence and crystal structure provided to us by our collaborator Dr. Uli Schwarz-Linek prior to the publication of the crystal structure (PDB 2X5P):

```
1          11         21         31         41         51
VDTLSGLSSE QGQSGDMTIE EDSATHIKFS KRDIDGKELA GATMELRDSS GKTISTWISD
61         71         81         91         101        111
GQVKDFYLMP GKYTFVETAA PDGYEVATAI TFTVNEQGQV TVNGKATKGD AHIVMVDA
```

CnaB2 is a domain from the fibronectin-binding protein FbaB from *S. pyogenes* (Hagan et al., 2010). The pDEST14-CnaB2 (wild-type CnaB2 construct) was a kind gift of Dr. Uli Schwarz-Linek (Oke et al., 2010). To construct Cna110 (CnaB2 residues 1-109), a stop codon was introduced at residue 110 in pDEST14-CnaB2 using the primer 5’–GGCAAAGCAACTAAAGGTTAAGCTCATATTGTCATGGTTG–3’ and its reverse complement. To generate Cna114 (CnaB2 residues 1-113), a stop codon was introduced at residue 114 in pDEST14-CnaB2 using the primer 5’– CTAAAGGTGACGCTCATATTTAAATGGTTGATGCTTGAGGATCC–3’ and its reverse complement. To create versions of Cna110 and Cna114 unable to form an isopeptide bond, the catalytic residue Glu^{77} was

mutated to Gln to make Cna110 E77Q and Cna114 E77Q, using the primer 5'–GGAAAATATACATTTGTCCAAACCGCAGCACCAGACG–3' and its reverse complement to mutate the pDEST14-Cna110 and pDEST14-Cna114 templates. To crosslink CnaB2, Cna110, Cna110 E77Q, and Cna114 to magnetic beads, a Cys residue was introduced at the N-terminal end of the constructs between the 6X His tags and the beginning of the CnaB2 coding sequences (Fig. 5.2a). To generate the Cys-CnaB2, Cys-Cna110, Cys-Cna110 E77Q, and Cys-Cna114 constructs, the primer 5'– CACCATCA CCATCACGATTGCGACATCCCAACGACCGAAAACC–3' and its reverse complement were used to mutate a Tyr residue to a Cys in the templates pDEST14-CnaB2, pDEST14-Cna110, pDEST14-Cna110 E77Q, and pDEST14-Cna114. To construct Cna114 I34D, Ile^{34} was mutated to Asp in pDEST14-Cna114 using primer 5'– CCATATTAAATTCTCA AAACGTGATGATGACGGCAAAGAGTTAGCTGG–3' and its reverse complement. To construct Cna114 I34E, Ile^{34} was mutated to Glu in pDEST14-Cna114 using the primer 5'– CCCATATTAAATTCTCAAAACGTGATGAGGACGGCAAAGAGTTAGC–3' and its reverse complement. To generate Cna114 I34N, Ile^{34} was mutated to Asn in pDEST14-Cna114 using primer 5'–CCATATTAAATTCTCAAAACGTGATAATGACGGCAAAGAG TTA–3' and its reverse complement. To make Cna114 I34R, Ile^{34} was mutated to Arg in pDEST14-Cna114 using primer 5'–CCATATTAAATTCTCAAAACGTGATAGGGACGGCA AAGAGTTAGC–3' and its reverse complement. To construct Cna114 I34S, Ile^{34} was mutated to Ser in pDEST14-Cna114 using primer 5'–CCATATTAAATTCTCAAAAC GTGATAGTGACGGCAAAGAGTTA–3' and its reverse complement. To generate Cna114 I34T, Ile^{34} was mutated to Thr in pDEST14-Cna114 using primer 5'–CCATATTAAATTC TCAAAACGTGATACTGACGGCAAAGAGTTAG–3' and its reverse complement. To construct Cna114 I58Y, Ile^{58} was mutated to Tyr in pDEST14-Cna114 using primer 5'–GGTAAAACTATTAGTACATGGTATTCAGATGGACAAGTGAAAGATTTC–3' and its reverse complement. Cna114 M69Y was created by mutating Met^{69} to Tyr in pDEST14-Cna114 using primer 5'– GATGGACAAGTGAAAGATTTCTACCTGTATCCAGGAAAAT ATACATTTGTCGAAACC–3' and its reverse complement. To generate SpyCatcher, Ile^{34}

was mutated to a Glu in pDEST14-Cna114 M69Y using primer 5'– CCCATATT AAATTCTCAAAACGTGATGAGGACGGCAAAGAGTTAGC–3' and its reverse complement. To develop a version of SpyCatcher unable to form an isopeptide bond, SpyCatcher E77Q was constructed by mutating Glu^{77} to Gln in pDEST14-SpyCatcher using primer 5'–GGAAAATATACATTTGTCCAAACCGCAGCACCAGACG–3' and its reverse complement. To generate Cna115, residue 115 was converted to a stop codon in pDEST14-CnaB2 using the primer 5'– CTAAAGGTGACGCTCATATTGTCTAAGTTGAT GCTTGAGGATCCGGAATTC–3' and its reverse complement. To construct Cna116, residue 116 was converted to a stop codon in pDEST14-CnaB2 using the primer 5'– GGTGACGCTCATATTGTCATGTAAGATGCTTGAGGATCCGGAATTCG–3' and its reverse complement.

The amino acid sequences of all of the CnaB2 based peptides made in this study are provided in Table 5.3. To construct a reactive peptide based on the CnaB2 wild-type sequence, PepA-MBP (CnaB2 residues 111-118) was generated using primers 5'– GGGGCATATGGGAGCCCACATCGTGATGGTGGACGCCGGTAGTGGTGAAAGTGGT AAAATCGAAGAAGGTAAA–3' (containing the peptide sequence and linker) and 5'– GGGGAAGCTT TTACGAGCTCGAATTAGTCTG–3' to amplify MBP from pMAL. The PCR product was then inserted in to the expression vector pET28a using *NdeI* and *HindIII* restriction sites, which gave a GSGESG linker between PepA and MBP. PepB-MBP was constructed similarly to PepA-MBP except the primers 5'– GGGGCATATGGGAGCCCAC ATCACCATGACCGACGCCGGTAGTGGTGAAAGTGGTAAAATCGAAGAAGGTAAA–3' and 5'– GGGGAAGCTTTTACGAGCTCGAATTAGTCTG–3' were used to amplify MBP from pMAL. To generate PepJ-MBP, where the residues RSG were introduced at the N-terminal end of PepA, the primer 5'– GCCGCGCGGCAGCCATATGGGACGATCGGG TGCCCACATCGTGATGGTGGA CG–3' and its reverse complement were used to alter pET28a-PepA-MBP. To introduce the residues GSR at the C-terminal end of PepA, PepK-MBP was generated by using the primer 5'– CCACATCGTGATGGTGGACGC CGGATCCCGAGGTAGTGGTGAAAGTGGTAAAATCG –3' and its reverse complement

to alter pET28a-PepA-MBP. To construct PepC-MBP (containing N-terminal RSG and C-terminal GSR additions to PepA), the primer 5'– GCCGCGCGGCAGCCATATGGGACG ATCGGGTGCCCACATCGTGATGGTGGACG–3' and its reverse complement were used to introduce the residues RSG in to pET28a-PepK-MBP. PepD-MBP was made by using the primer 5'– TGGGAGCCCACATCACGATGGTGGACGCCG–3' and its reverse complement to mutate pET28a-PepA-MBP. To construct PepE-MBP, the primer 5'– GCCCACATCGTGATGACGGACGCCGGTAGTGG–3' and its reverse complement were used to mutate pET28a-PepA-MBP. PepF-MBP was created using the primer 5'– TGGGAGCCCACATCAAGATGGTGGACGCCG–3' and its reverse complement to mutate pET28a-PepA-MBP. To generate PepG-MBP, the primer 5'– GCCCACATCGTGATGAA GGACGCCGGTAGTGG–3' and its reverse complement were used to mutate pET28a-PepA-MBP. PepH-MBP was constructed using the primer 5'– GGAGCCCACATCGAGAT GGTGGACGCC–3' and its reverse complement to mutate pET28a-PepA-MBP. To create PepI-MBP, the primer 5'– CCACATCGTGATGGAGGACGCCGGTAGTG–3' and its reverse complement were used to mutate pET28a-PepA-MBP. To generate PepL-MBP where the wild-type CnaB2 sequence was extended at the C-terminal end by 3 residues, residues YKP were introduced in to pET28a-PepA-MBP using the primer 5'– CCACATC GTGATGGTGGACGCCTACAAGCCGGGTAGTGGTGAAAGTGGTAAAATCG –3' and its reverse complement. PepM-MBP (CnaB2 residues 109-118) was constructed by introducing residues GD in to pET28a-PepA-MBP using primer 5'– CCGCGCGGC AGCCATATGGGAGGAGACGCCCACATCGTGATGGTGGACG–3' and its reverse complement. To make PepN-MBP, the primer 5'– AGCCATATGGGAGCCGAGATCGT GATGGTGGAC–3' and its reverse complement were used to mutate pET28a-PepA-MBP. To generate PepO-MBP, the primer 5'– GGGAGCCCACATCGTGATGTACGACGC CGGTAGTGGTGAAAG–3' and its reverse complement were used to mutate pET28a-PepA-MBP. PepP-MBP later renamed SpyTag-MBP was constructed by extending the natural CnaB2 sequence in PepA by 5 residues at its C-terminal end, where the primer 5'– GATGGTGGACGCCTACAAGCCGACGAAGGGTAGTGGTGAAAGTGGTAAAATCG–

3’ and its reverse complement were used to introduce the residues TK in to pET28a-PepL-MBP. To generate PepP-MBP D117A later renamed SpyTag-MBP D117A, a SpyTag variant unable to form an isopeptide bond, the primer 5’– CACATCGTGATGGT GGCCGCCTACAAGCCGAC–3’ and its reverse complement were used to mutate Asp^{117} to Ala in pET28a-SpyTag-MBP.

To display SpyTag on the surface of mammalian cells, SpyTag-CFP-TM (containing a GSGSG linker between SpyTag and CFP) was constructed using primers 5’–GGGGAGATCTGGCGGCGCCCACATCGTGATGGTGGACGCCTACAAGCCGACGA AGGGCTCGGGCTCCGGAATGGTGAGCAAGGGCGAG–3’ and 5’– CTCCCCGCGGC CGCTTGTATAGCTCGTCCATGCCGAGAGT–3’ to amplify CFP from pDisplay-Isopeptag-CFP-TM (Zakeri and Howarth, 2010). The PCR product was then inserted in to pDisplay (Invitrogen) using *BglII* and *SacII* restriction sites.

2.2. Protein expression and purification

2.2.1. General protein expression and purification procedures

Constructs cloned in pDEST14 were expressed using *E. coli* BL21 DE3 pLysS cells (Stratagene) and constructs cloned in pET28a were expressed using *E. coli* BL21 DE3 RIPL cells (Stratagene). All expression cultures were grown in Luria-Bertani (LB) containing 0.1 mg/mL ampicillin and 0.034 mg/mL chloramphenicol (pDEST14) or 0.05 mg/mL kanamycin (pET28a). Overnight cultures were diluted 100-fold in LB containing 0.8% glucose with the appropriate antibiotics, and subsequently grown at 37ºC to A_{600} 0.5, and then induced with 0.4 mM isopropyl β-D-thiogalactoside (IPTG) and incubated at 30°C for 4 hr. All proteins were purified using nickel affinity chromatography (except for CB ΔHis-MBP where amylose resin from New England Biolabs was used) with Ni-NTA resin (Qiagen).

All protein purifications were performed on ice or at 4°C. A typical protocol for purification of His tagged proteins using nickel affinity chromatography included re-suspending a bacterial pellet from an expression culture in 5 mL of Ni-NTA binding buffer

(50 mM Tris base, 300 mM NaCl, pH 7.8) containing 1 mM phenylmethylsulfonyl fluoride (PMSF), 1X EDTA-free protease mixed inhibitor cocktail (Roche), and 50 µg/mL DNase (for cultures expressed in *E. coli* BL21 DE3 pLysS). The re-suspended bacteria were then sonicated 4 times for 1 min at 50% pulse on an Ultrasonic Processor. The cell lysate was then centrifuged at 17,000 g in a micro-centrifuge for 10 min at 4°C, and the supernatant was added to 0.5 mL Ni-NTA resin (washed with Ni-NTA binding buffer) and the mixture was then incubated at 4°C for 30 min on a rotator. Subsequently, the mixture was centrifuged at 2,000 g for 2 min at 4°C. The resin was then re-suspended in 10 mL of Ni-NTA binding buffer containing 10 mM imidazole followed by centrifugation at 2,000 g for 2 min at 4°C, and this was repeated once. Subsequently, the resin was re-suspended in 1 mL of Ni-NTA binding buffer containing 30 mM imidazole and transferred to a 15 mL PolyPrep column (Bio-Rad) and the buffer was allowed to drain. Then 2 mL of Ni-NTA binding buffer containing 100 mM imidazole was added to elute His tagged proteins and 4 fraction of 0.5 mL were collected. Fractions were analyzed by measuring A_{280}, and fractions containing the highest concentration of purified protein were pooled and dialyzed. A similar protocol was used to purify CB ΔHis-MBP using amylose resin, except column buffer (50 mM Tris, 300 mM NaCl, 1 mM EDTA, pH 7.8) was used to re-suspend the bacterial pellet and to wash the resin following incubation with cell lysate, and then to elute the protein column buffer containing 10 mM maltose was used.

Purified proteins were then dialyzed at 4°C in to phosphate buffered saline (PBS) (Per L: 80.7 g NaCl, 2 g KCl, 18 g $Na_2HPO_4 \cdot 2H_2O$, 2.4 g KH_2PO_4, pH 7.4) or PBS containing 1 mM EDTA (if the constructs contained a Cys residue, since metal ions induce the oxidation of Cys residues (Ward et al., 1998)). The ExPASy ProtParam program was used to predict protein extinction coefficients which were used to determine protein concentrations from A_{280} values. Protein concentrations reported in Fig. 5.6 and chapter 6 were determined using a Microplate BCA Protein Assay Kit (Pierce). Protein expression yields were typically 10-20 mg per L of culture.

2.2.2. Expression of ^{15}N-Cna110

pDEST14-Cna110 was expressed using *E. coli* BL21 DE3 pLysS cells (Stratagene). The expression culture was prepared in 10 mL of Minimal media and incubated at 37°C for 16 hr. The recipe for 200 mL of Minimal media was: 150 mL autoclaved MilliQ H_2O, 40 mL 5X M9 medium (Per L: 33.4 g $Na_2HPO_4 \cdot 2H_2O$, 15.0 g KH_2PO_4, 2.5 g NaCl and autoclave), 0.2 mL autoclaved $CaCl_2$ (0.1M), 0.4 mL autoclaved $MgSO_4$ (1 M), 4 mL filtered sterilized 20% (w/v) glucose, 4 mL filter sterilized 8.5% (w/v) yeast nitrogenous base without amino acids or ammonium sulphate (YNB), 200 µL 100 mg/mL ampicillin, 400 µL 17 mg/mL chloramphenicol, 3 mL filter sterilized 10% (w/v) $^{15}NH_4Cl$ (a kind gift by Dr. Jason Schnell). The 10 mL overnight culture was then transferred to 190 mL Minimal media and incubated at 37°C for 16 hr. Subsequently, 4 baffled 2 L conical flasks were prepared with 450 mL 1X M9 medium and autoclaved and allowed to cool, then 0.5 mL autoclaved $CaCl_2$ (0.1M), 1 mL autoclaved $MgSO_4$ (1M), 10 mL filter sterilized 20% (w/v) glucose, 10 mL filter sterilized 8.5% (w/v) YNB, 500 µL 100 mg/mL ampicillin, 1 mL 17 mg/mL chloramphenicol, and 7.5 mL filter sterilized 10% (w/v) $^{15}NH_4Cl$ was added. The overnight culture was split between the 4 flasks and incubated at 37°C to A_{600} 0.5, and then induced with 0.4 mM IPTG and grown at 37°C for 16 hr. Cultures were then pooled and purified by nickel affinity chromatography using standard methods and dialyzed into 10 mM Na_2HPO_4 pH 7.4 at 4°C. This protocol yielded 14.5 mg of ^{15}N-Cna110.

2.3. SDS-PAGE

For SDS-PAGE analysis of proteins, all samples were heated to 95°C for 7 min in SDS loading buffer (Per 15 mL: 3.5 mL MilliQ H_2O, 3.5 mL 1 M Tris-HCl (tris-hydroxymethyl aminomethane) pH 6.8, 3.6 mL glycerol, 1.0 g SDS, 1.2 mg bromophenol blue) on a Bio-Rad C1000 thermal cycler to break all non-covalent interactions. Samples were then loaded on to polyacrylamide gels at indicated percentages and run at 200 V on an X-cel SureLock (Invitrogen). Subsequently, polyacrylamide gels were Coomassie

stained using either Coomassie brilliant blue followed by destaining in 60% MilliQ water, 30% methanol, and 10% acetic acid, or by Instant Blue (Triple Red Ltd.). For the quantification of band intensities, either a ChemiDoc XRS imager (Bio-Rad) and QuantityOne 4.6 software (Bio-Rad) or a Gel Doc™ XR imager (Bio-Rad) and Image Lab 3.0 software (Bio-Rad) was used.

2.4. Trypsin digestion reactions

Reactions to test the resistance of Spy0128 wild type (259 µM), Spy0128 E117A (173 µM), Spy0128 K179A (320 µM), and Spy0128 E117A K179A (290 µM) to trypsin digestion were performed by incubating 2 µL of each protein with either 8 µL of PBS pH 7.4 (– trypsin samples) or 8 µL of GIBCO cell culture 0.5 g/L trypsin EDTA (+ trypsin samples) (Invitrogen) for 15 min at 37°C. Reactions were subsequently analyzed on 18% SDS-PAGE with Coomassie staining.

2.5. Reconstitution reactions

Typically, reconstitution reactions were performed by mixing 1:1 ratios of peptides and protein partners at indicated concentrations, buffers, temperatures, and incubation times. Reactions were then stopped by heating to 95°C for 7 min in SDS loading buffer to break all non-covalent interactions followed by analysis using SDS-PAGE. For each figure shown, details regarding the reactions and incubation conditions are provided in the figure caption.

2.5.1. Reconstitution of split Spy0128 constructs

To analyze reconstitution between isopeptag-N and pilin-N, each protein was mixed at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25°C for 24 hr. Reconstitution between pilin-C -7 and CB-MBP or CB ΔHis-MBP was analyzed by mixing each protein at 25 µM in PBS pH 7.4 followed by incubation at 37°C for various times. To assess reconstitution between isopeptag-MBP and pilin-C, each protein was

mixed at 10 μM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25°C for 24 hr. To analyze the reconstitution efficiency of the various isopeptag-MBP and pilin-C variants (Fig. 3.8), each protein was mixed at 10 μM in PBS pH 7.4 and incubated at 25°C for 24 hr.

To investigate the speed, temperature-dependence, and pH-dependence of isopeptag-MBP and pilin-C reconstitution, 10 μM of each protein was mixed in 40 mM Na_2HPO_4 – 20 mM citric acid at the indicated pH and for the indicated amount of time. To analyze the ratio-dependence of reconstitution, the concentration of one partner was maintained at 10 μM while the concentration of the other partner was 0, 1, 2.5, 5, 10, 20 μM. To determine the concentration-dependence of the isopeptag-MBP and pilin-C reaction, the proteins were each mixed at equal concentrations of 1, 5, or 10 μM. To determine isopeptag-MBP and pilin-C reconstitution efficiency in various buffers and detergents, each protein was mixed at 10 μM in either PBS pH 7.4, 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.4, 50 mM Tris pH 7.4, or 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazine ethanesulfonate) pH 7.4, and also with or without 1% Triton X-100 or 0.5% Nonidet P-40 (NP40). 1 M NaOH or 1 M HCl was used to adjust the pH of all buffers, except for HEPES where 1 M KOH was used to determine whether isopeptide bond formation was dependent on the presence of Na^+. To stop all reactions, SDS loading buffer was added followed by heating at 95°C for 7 min to break all non-covalent interactions. Incubations at 4°C were placed in the refrigerator, while 25 and 37°C incubations were performed in a Bio-Rad C1000 thermal cycler, which was always used with a heated lid to prevent sample evaporation. To quantify the percentage of reconstitution between isopeptag-MBP and pilin-C, the density of the band corresponding to the isopeptag-MBP:pilin-C covalent complex was divided by the sum of the density of all the bands in the lane, and then multiplied by 100. To quantify the percentage of isopeptag-MBP or pilin-C reacted, the percentage reduction in the density of the corresponding band was determined relative to controls that did not contain the other reacting partner.

To demonstrate reconstitution of isopeptag-MBP and pilin-C inside bacteria, *E. coli* BL21 DE3 RIPL cells containing monocistronic or bicistronic expression vectors were grown at 37°C until A_{600} of 0.5, and then induced with 0.4 mM IPTG and incubated at 30°C for 4 hr. Samples from the expression cultures were then heated to 95°C for 7 min in SDS loading buffer, and then analyzed by SDS-PAGE with Coomassie staining.

2.5.2. Reconstitution of split ACE19 constructs

The split ACE19 constructs ACEA-MBP, ACE44, and ACE50 were mixed at 10 µM each in PBS pH 7.4 and incubated for 24 hr at 25°C. As a positive control for reconstitution Cna114 and PepA-MBP were used, and as a negative control Cna114 E77Q and PepA-MBP were used. Subsequently, samples were heated to 95°C for 7 min in SDS loading buffer to break all non-covalent interactions followed by analysis on a 14% SDS-PAGE with Coomassie staining.

2.5.3. Reconstitution of Cna peptide with Cna110

Cna peptide (RSGAHIVMVDAGSR) was synthesized by Insight Biotechnology Ltd and provided as a lyophilized white powder with >98% purity. A 10 mM stock solution of Cna peptide was made by dissolving it in *N,N*-dimethylformamide (DMF) and this stock was stored at 4°C. For testing reconstitution, 10 µM of Cna110 was mixed with 40 µM of Cna peptide in 10 mM Na_2HPO_4 pH 6.0 (the pH that was used in NMR spectroscopy) and the reaction was incubated at 25°C. Samples were taken at various time points and heated to 95°C for 7 min in SDS loading buffer and analyzed on 18% SDS-PAGE.

2.5.4. Reconstitution of ^{15}N-Cna110

Cna110 and ^{15}N-Cna110 were mixed with PepA-MBP each at 10 µM in 20 mM Na_2HPO_4 pH 7.4 and incubated at 25°C in a Bio-Rad C1000 thermal cycler for 3 hr. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining.

2.5.5. Reconstitution of split CnaB2 constructs

To analyze the reconstitution efficiency of Cna110 and Cna114 with PepA-MBP and PepB-MBP, proteins were mixed at 10 µM each in PBS pH 7.4 and incubated at 25°C for 24 hr. To perform a time course reaction of Cna110 and PepA-MBP, each protein was mixed at 10 µM in 20 mM Na_2HPO_4 pH 7.4 at 25°C and samples were removed at various time points for SDS-PAGE analysis. To analyze the temperature-dependence of the Cna110 and PepA-MBP reaction, proteins were mixed at 10 µM in 20 mM Na_2HPO_4 pH 7.4 and incubated at 4°C (in the refrigerator) or 25 and 37°C (in a Bio-Rad C1000 thermal cycler) for 1 hr.

To analyze the reconstitution efficiency of the various Cna114 mutants depicted in Table 5.2, the Cna114 mutants were incubated with either PepA-MBP or PepL-MBP each at 10 µM in PBS pH 7.5 for various time points, followed by SDS-PAGE analysis. The survey of SpyCatcher reactivity against the series of covalently reacting peptides (Table 5.3) was performed by mixing each protein at 10 µM in PBS pH 7.4 followed by incubation at 25°C for 30 min and SDS-PAGE analysis. Reconstitution efficiency of SpyTag-MBP with Cna114, Cna114 I34E, Cna114 M69Y, and SpyCatcher was analyzed by mixing each protein at 10 µM in PBS pH 7.4 at 25°C for 5 and 30 min followed by SDS-PAGE analysis.

Reconstitution between the optimized SpyTag-MBP and SpyCatcher pair was performed by mixing each protein at 10 µM in PBS pH 7.4 and incubation at 25°C for 3 hr. Reactions for analyzing the speed, temperature-dependence, and pH-dependence of the SpyTag-MBP:SpyCatcher pair, and the cross-reactivity of the SpyTag-MBP:SpyCatcher and isopeptag-MBP:pilin-C pairs were performed by mixing 10 µM of each protein in 40 mM Na_2HPO_4 – 20 mM citric acid at the indicated temperature and pH, and for the indicated time. To calculate the reaction rate for the SpyTag-MBP and SpyCatcher reaction, the data from the first 5 min of the time course reaction shown in Fig. 6.3b (where the reaction proceeded linearly) was used. A plot of 1/[SpyCatcher] (μM^{-1}) (amount of free SpyCatcher remaining in the mixture) versus time (min) was made. The slope of

the resulting trendline represented the second order reaction rate which was then converted to $M^{-1}s^{-1}$ (Price et al., 2001). To analyze the effect of various buffers on the reconstitution efficiency of SpyTag-MBP and SpyCatcher, each protein was mixed at 10 µM in either PBS pH 7.4, 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0, 50 mM Tris pH 7.0, or 50 mM HEPES pH 7.0 and incubated at 25°C for 1 and 10 min. To determine the ratio-dependence of SpyTag-MBP and SpyCatcher reconstitution, reactions were performed in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 for 3 hr at 25°C, where one reactant was maintained at 10 µM while the concentration of the other reactant was 0, 1, 2.5, 5, 10, and 20 µM. For analyzing the effect of detergent on SpyTag-MBP and SpyCatcher reconstitution, reactants were mixed at 10 µM each for 3 hr at 25°C in either PBS pH 7.4, or PBS pH 7.4 containing 1% Triton-X 100, 1% Tween20, or 0.5% NP40. Concentration-dependence of the SpyTag-MBP and SpyCatcher reaction was analyzed by incubating the reactants at equal concentrations of 1, 5, or 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 at 25°C for various times. All quantified reactions were performed in triplicate. To stop reactions, SDS loading buffer was added to the mixtures followed by heating to 95°C for 7 min. For determining temperature-dependence of SpyTag-MBP and SpyCatcher reconstitution, reactions were incubated in a Bio-Rad C1000 thermal cycler at 4, 25, and 37°C. Calculation of percent reconstitution was performed by dividing the intensity of the band for the SpyTag-MBP:SpyCatcher covalent complex by the intensity of all the bands in the lane, then multiplying by 100. To calculate the depletion of either SpyTag-MBP or SpyCatcher, the reduction in intensity of the band for each protein was monitored relative to a control lacking the reactive partner.

2.6. Mass spectrometry

2.6.1. Analysis of Spy0128

Purified Spy0128 was diluted to 20 µM in PBS pH 7.4 and analyzed on an electrospray ionization Quadrupole-Time-of-flight Micromass spectrometer (Waters, Manchester, UK) coupled to a High Performance Liquid Chromatography (HPLC) with an

autosampler (Dionex, Camberley, UK) at the Oxford Protein Production Facility (Nettleship et al., 2008).

2.6.2. Analysis of split Spy0128 constructs

Mass spectrometry experiments were performed on a Micromass LCT time-of-flight electrospray ionization mass spectrometer (Micromass, UK). For all reactions, proteins were mixed at 100 µM in PBS pH 7.4 and incubated at 4°C for 16 hr. Subsequently, reactions were dialyzed using 0.025 µm VSWP Millipore membrane filters (VSWP 01300) against MilliQ H_2O at 25°C for 1 hr. Following dialysis, reactions were further purified using Millipore ZipTip pipette tips (ZTC04S096) to remove residual salt. Maximum Entropy processing (MassLynx V4.00.00 software) was used to convert the m/z spectrums to a molecular mass profile. The ExPASy ProtParam program was used for the prediction of protein molecular masses based on the protein sequence.

2.6.3. Analysis of ACE19 and split CnaB2 constructs

To demonstrate that SpyCatcher interacted with Cna peptide through an isopeptide bond, 100 µM of SpyCatcher was mixed with 200 µM of Cna peptide and the reaction was incubated at 4°C for 16 hr. Subsequently, the reaction was dialyzed against MilliQ H_2O using 15.5 mm diameter dialysis tubing (BioDesign) 3 times for 2 hr at 4°C. ACE19 at 100 µM in PBS pH 7.4 was also dialyzed against MilliQ H_2O using 15.5 mm diameter dialysis tubing (BioDesign) 3 times for 2 hr at 4°C. Mass spectrometry and analysis was performed as above.

2.6.4. Analysis of ^{15}N-Cna110 and Cna peptide

Note: in this experiment (Fig. 4.12b) I prepared the ^{15}N-Cna110 protein and sent the protein along with Cna peptide to our collaborator (Dr. Uli Schwarz-Linek and his student Benjamin Schomburg). Dr. Uli Schwarz-Linek and Benjamin Schomburg

conducted the experiment at the St Andrews BSRC Mass Spectrometry and Proteomics Facility.

A ^{15}N-Cna110 and Cna peptide reaction mixture (20 µL reaction volume at 5 x 10^{-6} M) was de-salted using a XTerra MS C8 2.1 x 10 mm column. The sample was then eluted using an increasing acetonitrile concentration (2% acetonitrile, 98% aqueous, 1% formic acid, to 98% acetonitrile, 2% aqueous, 1% formic acid). The de-salted sample was then analyzed on an electrospray ionisation mass spectrometer (LCT, Micromass, Manchester, U.K.). Maximum Entropy processing was used to convert the m/z spectrum to a molecular mass profile.

2.7. SPASM searches

To find previously unrecognized intramolecular isopeptide bonds in proteins deposited in the Protein Data Bank (PDB), the SPASM software (Kleywegt, 1999) from the Uppsala Software Factory (http://xray.bmc.uu.se/usf/) was used to perform structural similarity searches using default software settings. Initially, the residues Lys^{36}, Asn^{168}, and Glu^{117} from the N-domain isopeptide bond in Spy0128 (PDB 3B2M) were used as the structural template with the allowed substitutions of Asp for Asn and Gln for Glu, but previously unrecognized intramolecular isopeptide bond containing proteins were not detected. Subsequently, the residues Lys^{176}, Asn^{293}, and Asp^{209} from a collagen binding domain in *Staphylococcus aureus* (PDB 2F68) were used as the template with the allowed substitution of Asp for Asn and Glu for Asp, which resulted in the identification of ACE19 (PDB 2OKM) as containing a previously unrecognized intramolecular isopeptide bond.

2.8. NMR spectrometry

Note: in this experiment I prepared the ^{15}N-Cna110 protein and sent the protein along with Cna peptide to our collaborator (Dr. Uli Schwarz-Linek and his student

Benjamin Schomburg). Dr. Uli Schwarz-Linek and Benjamin Schomburg performed the $^{1}H,^{15}N$-HSQC spectrometry experiment.

A 500 MHz Bruker DRX500 spectrometer equipped with a Z-shielded 54 mm gradient triple-resonance probe was used to conduct $^{1}H,^{15}N$-HSQC NMR experiments at 25°C. The Topspin 3.0 software (Bruker) was used to conduct the NMR experiment and analyze the data.

To prepare ^{15}N-Cna110 for NMR spectrometry, the N-terminal 6X His tag was removed from the recombinant protein by incubating with tobacco etch virus (TEV) protease for 16 hr. ^{15}N-Cna110 was then purified via reversed phase-HPLC (C4 matrix, acetonitrile gradient) and the fractions collected were freeze-dried. Subsequently, 10 mM Na_2HPO_4 was used to re-suspend freeze-dried protein fractions followed by the addition of 5% D_2O. The pH of the dissolved ^{15}N-Cna110 was adjusted to 6.0 using 1 M and 0.1 M NaOH. Based on A_{280} measurements (extinction coefficient of 9,970 $M^{-1}cm^{-1}$), the purified ^{15}N-Cna110 protein had a concentration of 0.18 mM. ^{15}N-Cna110 was then used to record a HSQC spectrum (WATERGATE water suppression, 8 scans, spectral resolution 9.76 and 11.08 Hz/point in the H and the N dimension, respectively). To assess reconstitution between Cna peptide and ^{15}N-Cna110, 0.3 mg of Cna peptide was added (pH of the resulting sample was re-adjusted) and another HSQC spectrum was recorded. The amount of time elapsed from Cna peptide addition and data acquisition was 18 min. The final concentration of Cna peptide was ~0.3 mM, which was calculated based on a final reaction volume of 0.6 mL, the weight of Cna peptide added, and 15% contribution of H_2O to the peptide weight.

2.9. Mammalian cell culture and cellular imaging

All mammalian cell culture experiments were performed using HeLa cells, which were grown in Dulbecco's modified eagle's medium (DMEM) containing 10% Fetal Calf Serum (PAA Labs), 50 U/mL penicillin, and 50 µg/mL streptomycin.

To display isopeptag on the surface of HeLa cells, 0.3 µg of pDisplay-isopeptag-CFP-TM was transfected in to HeLa cells in a well of a 48-well plate using 0.75 µL Lipofectamine 2000 (Invitrogen) based on manufacturer recommended procedures. Briefly, for each reaction 0.3 µg of pDisplay-isopeptag-CFP-TM and 0.75 µL Lipofectamine 2000 were each added separately to 25 µL of DMEM containing 50 U/mL penicillin and 50 µg/mL streptomycin, and the samples were mixed and incubated for 5 min at 25°C. The samples were then pooled and incubated at 25°C for 20 min. The HeLa cell media was then changed and 250 µL of DMEM containing 50 U/mL penicillin and 50 µg/mL streptomycin was added to each well, and then the transfection mixture was added to the cells. Cells were incubated at 37°C for 4 hr and then the cell media was removed and 0.5 mL of DMEM containing 10% Fetal Calf Serum, 50 U/mL penicillin, and 50 µg/mL streptomycin was added to each well. Subsequently, one day after transfection 5 µM of either pilin-C K179A or pilin-C in 300 µL of DMEM containing 10% Fetal Calf Serum, 50 U/mL penicillin, and 50 µg/mL streptomycin was added to the cells, and the cells were incubated for a further 4 hr at 37°C. Cells were then prepared for microscopic imaging as follows: cells were washed 3 times with PBS + 5 mM $MgCl_2$ (PBS-Mg) at 4°C, then 100 µL of 11 µg/mL penta-His antibody (Qiagen) in PBS-Mg with 1% BSA was added to the cells which were then incubated at 4°C for 10 min, cells were again washed 3 times with PBS-Mg at 4°C, then 100 µL of 20 µg/mL Alexa Fluor 555-anti-mouse antibody (Invitrogen) in PBS-Mg with 1% BSA was added and the cells were incubated at 4°C for 10 min, and the cells were again washed 3 times with PBS-Mg at 4°C. The cells were then fixed using 100 µL of fixing solution (PBS containing 4% formaldehyde, 4% sucrose, and 5 mM $MgCl_2$) for 15 min at 4°C, and washed 3 times with PBS-Mg.

To assess the display of SpyTag on the surface of HeLa cells, either 0.4 µg of pDisplay-isopeptag-CFP-TM or pDisplay-SpyTag-CFP-TM were transfected in to HeLa cells in a well of a 48-well plate using 1.0 µL Lipofectamine 2000 (Invitrogen) as described above. Subsequently, one day after transfection 5 µM of Cna114 I34E in 100 µL of DMEM containing 10% Fetal Calf Serum, 50 U/mL penicillin, and 50 µg/mL streptomycin was

added to the cells, which were then incubated for 5 min (SpyTag-CFP-TM) or 15 min (isopeptag-CFP-TM) at 25°C. Cells were then prepared for microscopic imaging as follows: cells were washed 3 times with PBS-Mg, then 100 µL of 33 µg/mL anti-His antibody (Abcam) in PBS-Mg with 1% BSA was added to the cells which were then incubated at 4°C for 10 min, cells were again washed 3 times with PBS-Mg, then 100 µL of 20 µg/mL Alexa Fluor 555-anti-mouse antibody (Invitrogen) in PBS-Mg with 1% BSA was added and the cells were incubated at 4°C for 10 min, and the cells were again washed 3 times with PBS-Mg. Cells were then stored on ice and imaged.

All microscopic imaging experiments were performed using a wide-field DeltaVision Core fluorescent microscope (AppliedPrecision) with a 40X oil-immersion lens. SoftWoRx 3.6.2 software was used to collect and analyze images of ECFP (436DF20 excitation, 480DF40 emission, Chroma 86002v1 dichroic) and Alexa Fluor 555 (540D420 excitation, 600DF50 emission, Chroma 84100bs polychroic). Images were all background corrected and exposure times of 0.1-1.0 s were typically used.

2.10. Phage display

2.10.1. Tethering of CnaB2 constructs to magnetic beads and ELISA plates

Dynabeads® M-270 Amine (Invitrogen) magnetic beads were coated with Sulfo-LC-SPDP (Sulfosuccinimidyl 6-(3'-[2-pyridyldithio]-propionamido)hexanoate) (Thermo Pierce) cross-linker using manufacturer recommended procedures. Briefly, 100 µL of magnetic beads were washed twice with PBS pH 7.5 containing 1/10 volume of freshly made 1 M $NaHCO_3$ pH 8.3, where a MagRack6 magnetic rack (GE Healthcare) was used to capture magnetic beads. The magnetic beads were then re-suspended with 89 µL of PBS pH 7.5 containing 1/10 volume of freshly made 1 M $NaHCO_3$ pH 8.3, and then 11 µL of sulfo-LC-SPDP dissolved in PBS pH 7.5 at 2 mg/mL was added. The magnetic beads were then incubated at 25°C for 1 hr (all incubation were stored on a shaker at 1400 rpm to prevent the beads from settling), then washed twice with PBS pH 7.5 containing 1/10 volume of freshly made 1 M $NaHCO_3$ pH 8.3. The activated beads were then coupled to

CnaB2 constructs modified to contain a N-terminal Cys residue by incubated the beads for 16 hr with an excess of protein under manufacturer recommended conditions. Briefly, the activated magnetic beads were washed twice with PBS pH 7.5 and then re-suspended in 70 uL PBS pH 7.5. Subsequently, 30 µL of Cys containing CnaB2 constructs (at ~250-500 µM stock concentrations) were added to the beads to provide reactive proteins at large excess. Subsequently, the magnetic beads were incubated at 25°C for 16 hr, and then reduced glutathione was added to the beads at a final concentration of 5 mM (to oxidize free Cys residues) and the magnetic beads were further incubated for 1 hr at 25°C. The protein conjugated magnetic beads were then washed 4 times with PBS pH 7.5 + 0.5% bovine serum albumin (BSA) + 0.1% Triton-X 100, and then re-suspended in 200 µL of PBS pH 7.5 + 3% BSA + 0.05% sodium azide followed by a 1 hr incubation at 25°C and storage at 4°C. To evaluate the coupling reaction, the release of pyridine-2-thione was detected by measuring absorbance at 343 nm on a NanoDrop ND-1000 (Thermo Scientific).

To tether Cys containing CnaB2 constructs to wells of 96-well polystyrene high binding enzyme-linked immunosorbent assay (ELISA) plates (Corning), proteins were incubated at 10 µM in 100 µL of PBS pH 7.4 per well for 16 hr at 4°C. The wells were then washed 3 times with Tris-Buffered Saline (TBS: 50 mM Tris-HCl + 150 mM NaCl) + 0.1% Tween-20 pH 7.5 for 2 min each time and then blocked with 200 µL of PBS pH 7.4 + 3% BSA + 5 mM reduced glutathione for 1 hr at 25°C on a rocker.

2.10.2. Phage display panning experiments

All phage display experiments were performed using the Ph.D.^{TM}-12 Phage Display Library Kit (New England Biolabs). Panning experiments were conducted at 100 µL volumes against CnaB2 constructs tethered to either magnetic beads or ELISA plates. All experiments were performed using manufacturer recommended procedures and reactions involving magnetic beads were performed at 1400 rpm on a rocker to prevent the beads from settling in the reaction vessels (New England Biolabs, 2009). Briefly,

coated magnetic beads or ELISA plate wells were washed 3 times with 100 µL of TBS pH 7.5 + 0.1% Tween20, then 99 µL of TBS pH 7.5 + 0.1% Tween20 and 1 µL of the phage library was added to the magnetic beads/wells. The reactions were then incubated at 25°C for various times as outlined in Table 5.1. The panning experiments were customized for isolating covalently binding peptides by performing wash steps in the panning protocol under very stringent conditions as outlined in Table 5.1. To elute bound phage following stringent wash steps, beads were incubated with either 100 µL of PBS pH 7.5 + 50 mM dithiothreitol (DTT) for 2 hr at 25°C or 100 uL of TEV protease at 13 µM in 50 mM Tris-HCl pH 8.0 + 0.5 mM EDTA for 2 hr at 34°C (Haering et al., 2008). The plasmid encoding for recombinant TEV protease was a kind gift from Dr. Kim Nasmyth (Haering et al., 2008), and it was expressed and purified using nickel affinity chromatography as outlined in section 2.2.1. For panning experiments performed on ELISA plates, bound phage were not eluted but were titred directly in the wells. Briefly, to titre phage a 10 mL *E. coli* ER2738 (New England Biolabs) culture was grown in LB + 20 µg/mL tetracycline until A_{600} of 0.5. Subsequently, the eluate was serially diluted 10-10^4 fold in 15 µL LB (final volume) and 10 µL of each dilution was added to 200 µL of the *E. coli* ER2738 culture and the cells were mixed and incubated at 25°C for 5 min. Then 3 mL of top agar (per L: 10 g bacto-tryptone, 5 g yeast extract, 5 g NaCl, 7 g agar) at 45°C was added to the cells, which were then vortexed and poured on to LB + 20 µg/mL tetracycline + IPTG + 5-bromo-4-chloro-3-indolyl β-D-galactoside (Xgal) plates (Per L of LB: 15 g agar, 20 mg tetracycline, 50 mg IPTG, 40 mg Xgal). The plates were allowed to cool and then incubated at 37°C for 16 hr. Subsequently, phage plaques were counted and titred. After each round of panning, individual phage were picked from titration plates and amplified using manufacturer recommended procedures (New England Biolabs, 2009) and screened using Western blot analysis. To amplify phage, *E. coli* ER2738 was grown in LB + 20 µg/mL tetracycline at 37°C for 16 hr. The culture was then diluted 1:100 and individual phage plaques from the titration plates were each used to inoculate 2 mL of the diluted *E. coli* ER2738 culture, which were then incubated for 4.5 hr at 37°C. The

amplified phage cultures were then centrifuged at 17000 g for 3 min, and the supernatants (containing the phage) were removed and stored at 4°C.

2.10.3. Western blots for phage display screens

Amplified phage that were isolated from panning experiments were screened by Western blot analysis to determine if they covalently reacted with Cna110. A typical reaction would involve incubating 5 µL of Cna110 (151 µM stock concentration) with 10 µL of amplified phage stock (typically at 10^{13} pfu/mL) to ensure reactants were present in excess in order to enable detection of small amounts of covalent product. Reactions were then incubated for 6-16 hr at 25°C then heated to 95°C for 7 min in SDS loading buffer containing 50 mM DTT and run on 10% SDS-PAGE for 1 hr at 200 V in a X-cel SureLock. The gels were then transferred to a Trans-blot Transfer Medium Pure Nitrocellulose membrane (Bio-Rad) in Transfer buffer (For 1 L: 1.45 g Tris-base, 7.2 g glycine, 200 mL methanol, 800 mL MilliQ H_2O) for 60 min at 25 V in a X-cel SureLock. The membranes were then blocked in 20 mL of PBS pH 7.4 containing 2.5 g fat-free milk for 1 hr on a rocker. The membranes were then washed 3 times with PBS pH 7.4 + 0.05% Tween-20 and incubated with 1 mL of a 1:1000 dilution of anti-M13 pIII monoclonal antibody (New England Biolabs) in PBS pH 7.4 + 0.05% Tween-20 for 1 hr at 25°C. The membranes were then washed 4 times for 1 min with PBS pH 7.4 + 0.05% Tween-20 on a rocker and incubated with 15 mL of a 1:5000 dilution of anti-mouse IgG-horseradish peroxidase secondary antibody (Sigma) in PBS pH 7.4 + 0.05% Tween-20 for 40 min at 25°C on a rocker. The membranes were then washed 4 times for 5 min with PBS pH 7.4 + 0.05% Tween-20 on a rocker, then blotted dry and incubated with 2 mL of SuperSignal® West Pico (Thermo Pierce) for 5 min. The membranes were then blotted dry and allowed to develop on an Amersham Hyperfilm™ ECL film (GE Healthcare) for 1-10 min. Films were developed on a Compact X4 (Xograph Imaging System) using Devalex M solutions (Champion Photochemistry).

Chapter 3:

Dissection of Spy0128 to develop a spontaneously and covalently reconstituting peptide-protein pair

Parts of this chapter have been published in (Zakeri and Howarth, 2010), which is provided in the appendix.

Chapter 3: Dissection of Spy0128 to develop a spontaneously and covalently reconstituting peptide-protein pair

In 2007, the crystal structure of the major pilin subunit Spy0128 from *S. pyogenes* was determined and it showed that the protein contained two autocatalyzed intramolecular isopeptide bonds, one in each of the two domains of the protein (Fig. 1.6) (Kang et al., 2007). These bonds are strategically positioned to provide enhanced structural stability and protease resistance to the protein (Fig. 1.7) (Kang and Baker, 2009; Alegre-Cebollada et al., 2010). Based on the spontaneous nature of isopeptide bond formation, it was hypothesized that an intramolecular isopeptide bond containing protein could be split in to a small peptide fragment and a larger protein binding partner that could spontaneously interact via intermolecular isopeptide bond formation. The peptide and protein partner would each contain one residue involved in isopeptide bond formation, and when in solution they would be able to recognize and bind each other and form an isopeptide bond in a reaction catalyzed by a specific basic residue present in the binding partner. At the time of this investigation, Spy0128 was the only protein experimentally confirmed to contain spontaneously formed intramolecular isopeptide bonds. Therefore, it was decided to conduct proof of principle experiments for the development of a peptide-protein pair that could interact via isopeptide bond formation, by splitting the Spy0128 scaffold.

3.1. Recombinant Spy0128 was able to form two intramolecular isopeptide bonds

Prior to splitting Spy0128, it was important to determine whether recombinantly expressed Spy0128 was able to form two isopeptide bonds under the experimental conditions used in these investigations. Therefore, the *Spy0128* gene from *Streptococcus pyogenes* M1 GAS strain SF370 genomic DNA (constituting residues 18-308 that were used for structural determination by Kang et al. (Kang et al., 2007)) was cloned in to the expression vector pET28a (Novagen), which fused a N-terminal 6X His tag to the

construct. Subsequently, recombinant Spy0128 was expressed and purified using immobilized metal-affinity chromatography, and analyzed via liquid chromatography electrospray ionization mass spectrometry (Fig. 3.1b). Recombinant Spy0128 had an expected molecular mass of 34,324.0 Da based on its predicted sequence, however for each isopeptide bond that was formed a NH_3 molecule was released (Fig. 3.1a). Therefore, recombinant Spy0128 containing two intramolecular isopeptide bonds had an expected molecular mass of 34,290.0 Da (34,324.0 Da – 34 Da). Mass spectrometry of recombinant Spy0128 revealed a protein with a mass of 34,288 Da, which was in close agreement with the expected mass of recombinant Spy0128 containing two isopeptide bonds between Lys and Asn side chains (Fig. 3.1b). Therefore, the experimental conditions used favoured isopeptide bond formation and I proceeded to characterize the stability of Spy0128 by testing its resistance to trypsin digestion.

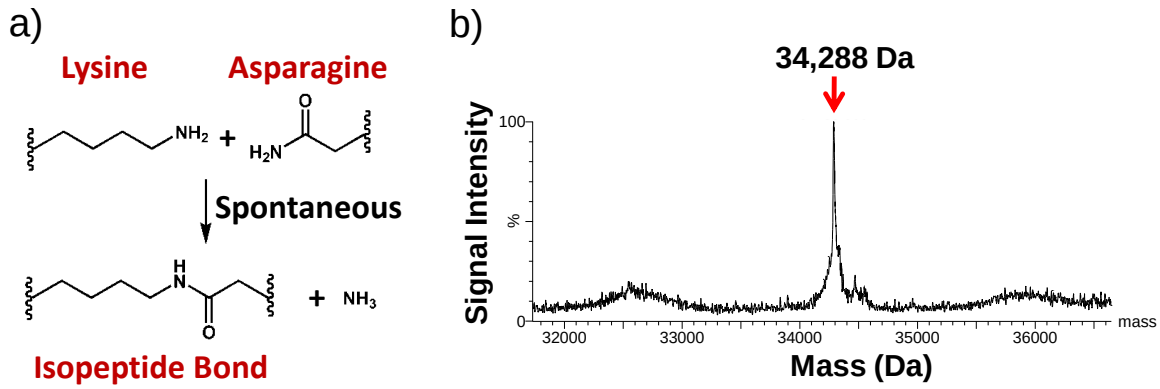


Figure 3.1. Recombinant Spy0128 formed two intramolecular isopeptide bonds. a) Isopeptide bond formation in Spy0128 occurs between Lys (unprotonated) and Asn side chains resulting in the release of a NH_3 molecule for each isopeptide bond formed. b) Liquid chromatography electrospray ionization mass spectrometry of recombinant Spy0128 revealed a protein with a mass of 34,288 Da. Based on its predicted sequence, recombinant Spy0128 had a molecular mass of 34,324.0 Da and each isopeptide bond released a NH_3 molecule with a mass of 17 Da. Therefore, Spy0128 containing two isopeptide bonds was predicted to have a mass of 34,290.0 Da.
Figure 3.1a was adapted from (Zakeri and Howarth, 2010).

3.2. Spy0128 was resistant to trypsin digestion

Having established that the recombinant Spy0128 protein contained two intramolecular isopeptide bonds, I next proceeded to investigate the contribution of each

bond to the resistance of Spy0128 to trypsin digestion. As illustrated in Fig. 1.6, Spy0128 contains one isopeptide bond in its N-terminal domain involving residues Lys^{36}, Asn^{168}, and Glu^{117}, and another isopeptide bond in its C-terminal domain involving residues Lys^{179}, Asn^{303}, and Glu^{258}. These isopeptide linkages form between residues situated at the beginning and end of the polypeptide sequence of each domain. Thus, when the protein is unfolded its structure would resemble a 'figure eight', constituting two linked circular chains (Fig. 3.2a). However, if the isopeptide forming residues were to be mutated, then the corresponding domain would be unable to form a circular linkage and it would have a linear structure when denatured (Fig. 3.2a). Therefore, if various isopeptide-deficient mutants of Spy0128 were analyzed on SDS-PAGE, it would be expected that the circular and linear polypeptides would have varying mobilities as they make their way through the gel matrix, where the circular polypeptide structures may migrate faster as a result of a more compact structure.

Using site-directed mutagenesis, I developed various isopeptide-deficient mutants of Spy0128 (Fig. 3.2a). I developed a mutant unable to form an isopeptide bond in the N-domain (Spy0128 E117A) by mutating the catalytic Glu^{117} residue, a mutant unable to form the C-domain isopeptide bond (Spy0128 K179A) by mutating the reactive Lys^{179} residue, and the isopeptide null double mutant Spy0128 E117A K179A. Interestingly, when these mutants were run on a SDS-PAGE gel, the E117A mutant migrated approximately the same distance as the wild-type construct, while the K179A mutant migrated a shorter distance followed by the E117A K179A double mutant (Fig. 3.2b). Since the E117A mutation did not alter the migration of Spy0128 on a SDS-PAGE gel, it is unclear why the E117A K179A double mutant migrated less than the K179A mutant (Fig. 3.2b).

I further went on to test the resistance of the various Spy0128 mutants to trypsin digestion. Proteins were incubated with either PBS pH 7.4 or trypsin for 15 min at 37°C, and then analyzed by SDS-PAGE (Fig. 3.2b). As expected, wild-type Spy0128 did not show any degradation in the presence of trypsin, but surprisingly neither did the K179A

mutant (Fig. 3.2b). On the other hand, both the E117A mutant and the E117A K179A double mutant were almost completely degraded within 15 min in the presence of trypsin, as can be observed with the loss of intensity of the bands corresponding to the Spy0128 variants and the formation of many smaller bands (Fig. 3.2b). One possible explanation would be that an isopeptide null C-domain is more stable than an isopeptide null N-domain. It has been shown that isopeptide-deficient mutants of Spy0128 were able to fold in to the correct protein conformation with an almost identical crystal structure when compared to wild-type Spy0128 (Kang and Baker, 2009). Yet 172 ± 33 pN of force was required to unfold the N-domain compared to 250 ± 96 pN for the C-domain (Alegre-Cebollada et al., 2010). This enhanced stability could be attributed to the fact that the β-strands that were isopeptide bonded together in the N-domain were also connected through six hydrogen bonds, while nine hydrogen bonds were present in the C-domain (Alegre-Cebollada et al., 2010). Therefore, the C-domain was better able to maintain its shape and resist trypsin digestion as seen with the Spy0128 K179A mutant (Fig. 3.2b). Alternatively, the E117A mutation may have been more destabilizing than the K179A mutation. It has been shown that in the N-domain isopeptide-deficient mutants E117A and N168A, the N168A mutant was more stable than the E117A mutant (Kang and Baker, 2009). This was because mutation of Glu^{117} caused a disturbance to the hydrophobic core of the protein due to the increased flexibility of the Lys^{36} and Asn^{168} side chains that were no longer isopeptide bonded (Kang and Baker, 2009). However, in the Asn^{168} mutant the side chain of Lys^{36} formed a salt bridge with the side chain of Glu^{117}, thereby reducing the disturbance introduced in to the core of Spy0128 as a result of the removal of the isopeptide bond (Kang and Baker, 2009). Consequently, the choice of the residue to mutate in order to prevent isopeptide bond formation was very important as it could affect protein stability.

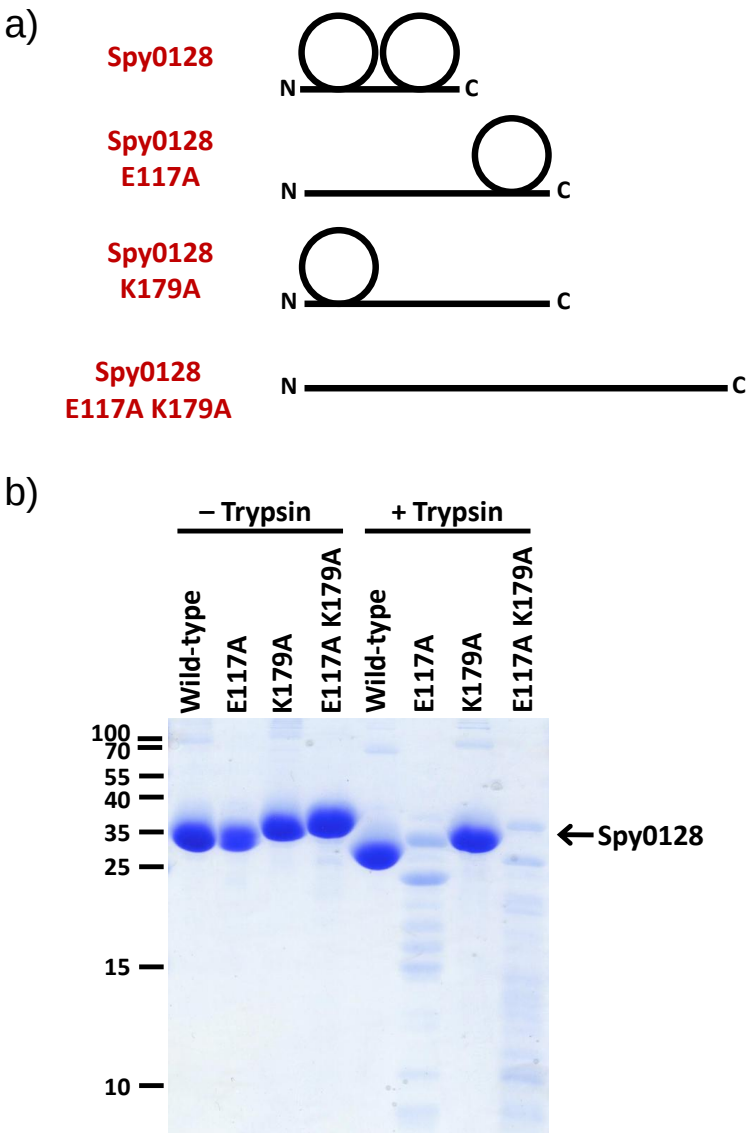


Figure 3.2. Trypsin digestion of Spy0128 and its isopeptide-deficient mutants. a) Schematic illustrating the unfolded polypeptide structure of Spy0128 and its isopeptide-deficient mutants. The presence of isopeptide bonds at the beginning and end of both the N- and C-domains form cyclic conformations in the unfolded polypeptide. b) Spy0128 and its isopeptide-deficient mutants were incubated with PBS pH 7.4 or trypsin for 15 min at 37°C, and subsequently analyzed on 18% SDS-PAGE with Coomassie staining. The further migration of wild-type and K179A Spy0128 in the + Trypsin sample compared to the – trypsin sample may be due to cleavage of the N-terminal 6X His tag by trypsin.

3.3. Splitting of Spy0128

3.3.1. Strategy for the development of covalent peptide-protein pairs

In order to develop peptide-protein pairs that could bind each other through spontaneous isopeptide bond formation, I had to split the Spy0128 protein. Some

concerns regarding splitting of the protein included the solubility of the subsequent fragments due to exposure of core hydrophobic regions to solution, which could lead to protein aggregation and precipitation. Also, the resulting fragments would likely be in an unfolded state and thus more prone to proteolytic digestion. Even if these concerns were overcome, there was still the requirement for the peptide and protein fragments to come together and fold in to the proper protein structure, and then for the three residues involved in isopeptide bond formation to be positioned in the correct orientation to facilitate linking of the two polypeptides through an isopeptide bond. Therefore, I had to be very selective in the positions that I chose to split Spy0128.

Inspection of the Spy0128 structure showed that it could be strategically split in two different regions (Fig. 3.3). As mentioned earlier, in both the N- and C-domains the isopeptide bonds link the first and last β-strands. Therefore, the best approach to making a short peptide tag and its corresponding binding partner was to simply remove either the N- or C-terminal β-strands (Fig. 3.3). This would create a short unstructured peptide that could potentially be used as a fusion tag. The length of the subsequent peptide was important as it had to be long enough to ensure binding specificity for the binding partner, but as short and unstructured as possible so that when fused to a target protein it would not affect protein function or solubility. Also, in developing the protein binding partner it would be ideal to minimize the number of residues removed from the Spy0128 scaffold. This may help the protein fragment to better fold to a structure resembling the wild-type protein, with the small region removed acting like a binding pocket for the peptide, and also to preserve the hydrophobic environment that is crucial for reaction around the isopeptide bond forming residues. Another consideration was the exact residue at which the protein was to be split. I believed that splitting of the protein should affect the overall protein fold as little as possible. Therefore, I concluded to split the protein in loop regions to have minimal structural impact.

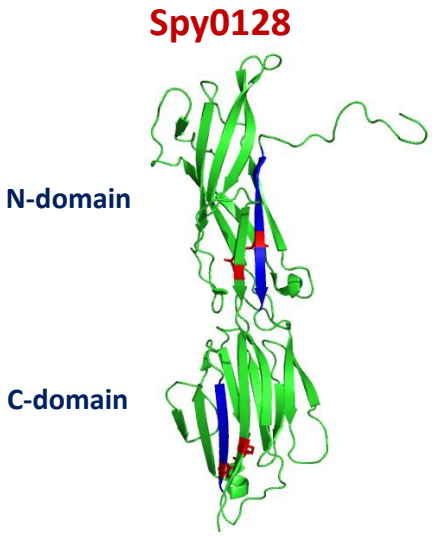


Figure 3.3. Strategic locations to split Spy0128. In both the N- and C-domains the first and last β-strands are linked by isopeptide bonds (red). The best strategy to make a covalently binding peptide-protein pair would be to split the protein in loop regions and remove either the first β-strand in the N-domain or the last β-strand in the C-domain (both shown in blue). Figure made from PDB 3B2M.

3.3.2. Splitting of Spy0128 in the N-terminal domain

To separate the isopeptide bond in the N-domain (Fig. 3.4a) and split the protein, I dissected Spy0128 in the loop region after the N-terminal β-strand (Fig. 3.4b). This created a 28 residue peptide that was named isopeptag-N (Spy0128 residues 18-45: ATTVHGETVVNGAKLTVTKNLDLVNSNA), and the corresponding protein binding partner that was named pilin-N (Spy0128 residues 46-308) (Fig. 3.4b). To aid in recombinant expression of isopeptag-N, the peptide was genetically linked at its N-terminal end to the maltose binding protein (MBP) with a six residue linker in between. This resulted in the reactive Lys^{36} residue to be positioned on isopeptag-N, while the reactive residues Asn^{168} and Glu^{117} were positioned on pilin-N. To test whether the peptide-protein pair was able to covalently reconstitute, purified MBP-isopeptag-N and pilin-N were mixed in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 at a concentration of 10 µM of each protein and then incubated at 25°C for 24 hr. Subsequently, the reactions were heated to 95°C for 7 min in SDS loading buffer to break all non-covalent interactions and then analyzed on SDS-PAGE (Fig. 3.4c). Reaction between MBP-isopeptag-N and pilin-N resulted in the

formation of a new product at ~80 kDa corresponding to a covalently linked MBP-isopeptag-N:pilin-N complex (Fig. 3.4c). However, when MBP-isopeptag-N was similarly incubated with a pilin-N mutant where the catalytic Glu^{117} was mutated to an Ala (pilin-N E117A) there was no new product formed (Fig. 3.4c). Therefore, isopeptag-N and pilin-N were able to spontaneously reconstitute in solution and form a stable isopeptide bond, thereby validating the hypothesis.

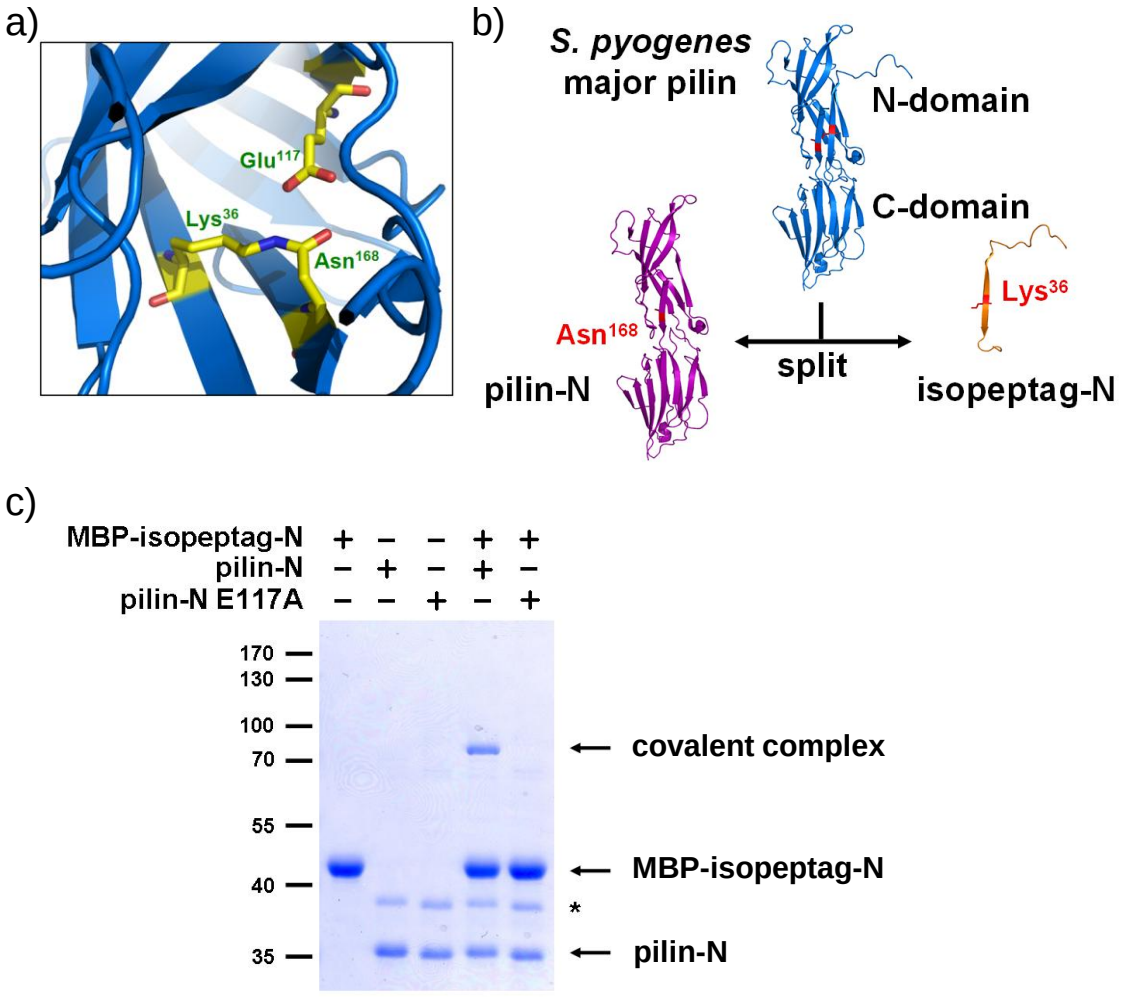


Figure 3.4. Spontaneous intermolecular isopeptide bond formation between MBP-isopeptag-N and pilin-N. a) Lys^{36} and Asn^{168} in the N-terminal domain of Spy0128 form an intramolecular isopeptide bond in a reaction catalyzed by Glu^{117} (PDB 3B2M). b) Spy0128 was split by removing the first β-strand in the N-terminal domain to make isopeptag-N (orange) and its corresponding protein binding partner pilin-N (purple). Residues involved in isopeptide bond formation are shown in red. c) To demonstrate covalent reconstitution between isopeptag-N and pilin-N, MBP-isopeptag-N and pilin-N or pilin-N E117A were mixed at 10 μM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25°C for 24 hr. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining. *Pilin-N expression

resulted in the formation of a side-product, which may be due to incomplete isopeptide bond formation in the C-terminal domain of pilin-N.
This figure was published in (Zakeri and Howarth, 2010).

3.3.3. Splitting of Spy0128 in the C-terminal domain

I next used the same approach to separate the isopeptide bond in the C-domain (Fig. 3.5a) and thereby split Spy0128 in the loop region before the C-terminal β-strand (Fig. 3.5b). This created a 16 amino acid peptide which was named CB (Spy0128 residues 293-308: TDKDMTITFTNKKDFE) and its corresponding protein partner that was called pilin-C -7 (Spy0128 residues 18-292) (Fig. 3.5b). The nomenclature for pilin-C -7 is based on optimization experiments conducted on the split pilin (Section 3.4.3), where the construct chosen for all future experiments was named pilin-C. To aid in recombinant expression of CB, MBP was fused to the C-terminal end of CB with a six residue linker in between. Therefore, the reactive Asn^{303} residue was positioned on CB and the reactive Lys^{179} and Glu^{258} residues were positioned on pilin-C -7 (Fig. 3.5b). The constructs were then expressed, purified, and tested for covalent reconstitution. CB-MBP and pilin-C -7 were mixed at a concentration of 25 μM of each protein in PBS pH 7.4 and incubated at 37°C. Samples were taken at various time points and analyzed on SDS-PAGE (Fig. 3.5c). Reaction between CB-MBP and pilin-C -7 produced a new product at ~80 kDa corresponding to a covalently linked CB-MBP:pilin-C -7 complex (Fig. 3.5c). Interestingly, the CB-MBP:pilin-C -7 pair reacted much more efficiently than the MBP-isopeptag-N:pilin-N pair (compared Fig. 3.4c and Fig. 3.5c). Moreover, the CB peptide was 12 residues shorter than the isopeptag-N peptide, thus making it more desirable for future use as a peptide fusion tag. Taking these factors in to consideration, I decided to conduct all further experiments using the CB:pilin-C -7 platform.

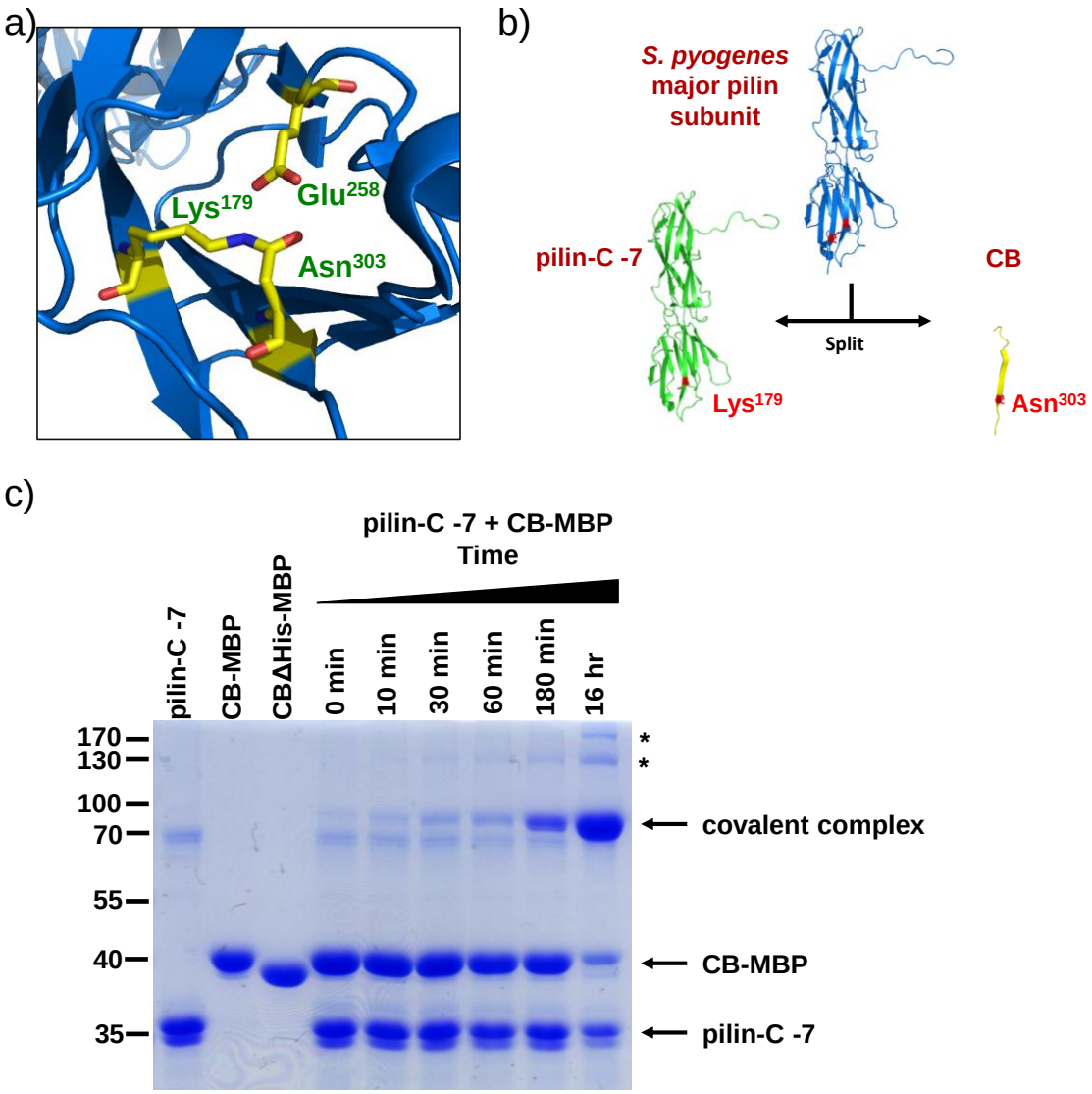


Figure 3.5. Spontaneous intermolecular isopeptide bond formation between CB-MBP and pilin-C -7. a) Lys^{179} and Asn^{303} in the C-terminal domain of Spy0128 form an intramolecular isopeptide bond in a reaction catalyzed by Glu^{258} (PDB 3B2M). b) Spy0128 was split by removing the last β-strand in the C-terminal domain to make CB (yellow) and its corresponding protein binding partner pilin-C -7 (green). Residues involved in isopeptide bond formation are shown in red. c) To demonstrate covalent reconstitution between CB and pilin-C -7, CB-MBP and pilin-C -7 were mixed each at 25 µM in PBS pH 7.4 and incubated for various times at 37°C. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 12% SDS-PAGE with Coomassie staining. CBΔHis-MBP lacks the N-terminal 6X His tag present in CB-MBP and was a control for an experiment discussed in section 3.4.1. *The CB-MBP and pilin-C -7 reaction resulted in the formation of side products which may have been a result of cross-reactions between multiple pilin-C -7 and CB-MBP molecules.
Figure 3.5a was published in (Zakeri and Howarth, 2010) and Figure 3.5b was adapted from (Zakeri and Howarth, 2010).

3.4. Optimization of split constructs

Having established that it was possible to develop a peptide-protein pair that could react spontaneously in solution via intermolecular isopeptide bond formation, I next proceeded to try to optimize the efficiency of reconstitution. Since the C-domain split constructs were more efficient at reconstitution than their N-domain counterparts, I proceeded to conduct all optimization experiments with the CB:pilin-C -7 pair.

3.4.1. Removal of the N-terminal 6X His tag from CB-MBP

The CB-MBP construct was cloned in to the expression vector pET28a, which fused a 6X His tag and a protease cleavage site at the 5' end of the inserted gene, resulting in the addition of 20 residues at the N-terminal end of CB. These extra residues may have interfered with the binding of the peptide to the protein partner. Therefore, I removed the extra residues introduced by the pET28a vector to make CBΔHis-MBP in order to assess whether display of CB directly at the N-terminal end of MBP had an effect on reconstitution efficiency. Subsequently, I mixed CBΔHis-MBP and pilin-C -7 at a concentration of 25 µM of each protein in PBS pH 7.4 and incubated the reaction at 37°C. At various time points samples were removed and analyzed on SDS-PAGE (Fig. 3.6). Surprisingly, CBΔHis-MBP was less efficient than CB-MBP at covalent reconstitution with pilin-C -7 (compare Fig. 3.5c to Fig. 3.6). This implied that the extra residues from the 6X His tag at the N-terminal end of CB in fact helped to position CB in a conformation that was slightly more favourable for reaction. The CB sequence has a cluster of uncharged polar and hydrophobic residues at its centre (CB sequence with core uncharged polar and hydrophobic residues shown in red: TDKD**MTITFT**NKKDFE), which may mediate hydrophobic interactions between the peptide and the surrounding environment including hydrophobic residues on the fusion protein MBP. The N-terminal 6X His tag may have increased the solubility of CB resulting in better display of the peptide in solution, and thus a more favourable conformation for reconstitution with pilin-C -7.

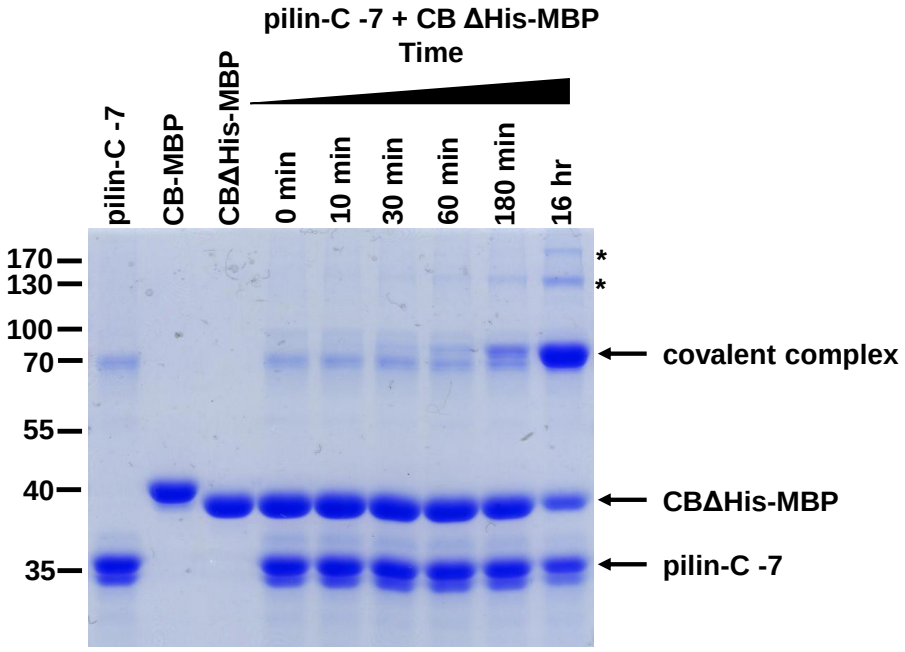


Figure 3.6. Reconstitution between CBΔHis-MBP and pilin-C -7. To assess covalent reconstitution between CBΔHis and pilin-C -7, CBΔHis-MBP and pilin-C -7 were mixed in PBS pH 7.4 each at 25 µM, and incubated for various times at 37°C. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 12% SDS-PAGE with Coomassie staining. CB-MBP is a control for an experiment discussed in section 3.3.3. * The CBΔHis-MBP and pilin-C -7 reaction resulted in the formation of side products which may have been a result of cross-reactions between multiple pilin-C -7 and CBΔHis-MBP molecules.

3.4.2. Rationale for the generation of the CB variant isopeptag

There is an exposed hydrophobic residue, Phe^{307}, At the C-terminal end of Spy0128 (Fig. 3.7). During pilus assembly when Spy0128 subunits are covalently linked in a head-to-tail orientation, Phe^{307} is packed between two Spy0128 subunits (Kang et al., 2007). However, when Spy0128 was dissected to make the CB:pilin-C -7 pair (Fig. 3.5b), Phe^{307} was positioned on the peptide CB. To increase the hydrophilic nature of CB, I mutated the exposed Phe^{307} residue to an Ala and termed the new peptide isopeptag (isopeptag sequence: TDKDMTITFTNKKDAE). Since, Phe^{307} is present in an unstructured region of Spy0128 (Fig. 3.7), it is likely not involved in the binding of CB to pilin-C -7. Moreover, a more hydrophilic peptide is likely to be better displayed in solution and less likely to adhere to surrounding hydrophobic surfaces. Therefore, even though the F307A mutation may not have an effect in the peptide-protein interaction, it is still

advantageous to have a more hydrophilic peptide for downstream application of this technology.

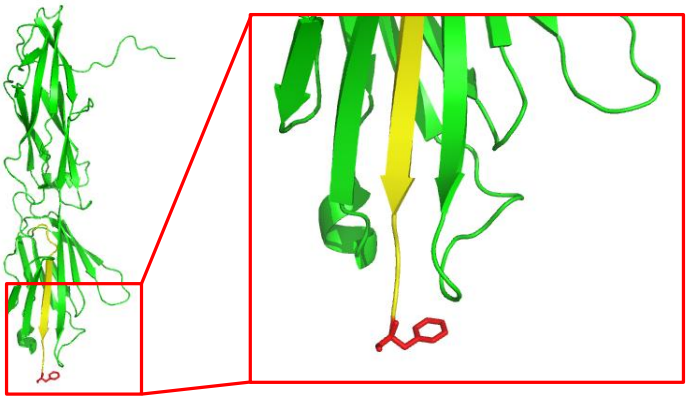


Figure 3.7. Phe^{307} is a hydrophobic residue in an unstructured region. To increase the solubility of the CB, Phe^{307} (red) was mutated to Ala to make the new peptide isopeptag (PDB 3B2M). CB is shown in yellow and pilin-C -7 is shown in green.

3.4.3. Rationale for optimization of the lengths of isopeptag and pilin-C -7

I also explored altering the sizes of the two interacting partners to see what effects this had on reconstitution efficiency. To create pilin-C -7, Spy0128 was split in a loop region preceding the C-terminal β-strand, thereby creating a void in the C-domain and exposing core hydrophobic regions to solution. This could potentially cause a destabilization of the protein or cause the protein to close the newly formed gap, and fold in such a way so as to reduce the exposed hydrophobic surface area. Thus, the short peptide sequence may have an even smaller surface area available for binding on the protein partner. As a result, I investigated incrementally increasing the length of pilin-C -7 at its C-terminal end and extending the protein from the site it was split in a flexible loop, in to the final β-strand in Spy0128 (Fig. 3.8a). This could help the fragmented protein to better fold in solution and obtain a more stable structure resembling the wild-type Spy0128, with an exposed cavity where the final β-strand was removed. Moreover, this could also help to reduce potential alternative conformations the fragmented protein may adopt and help to maintain a more rigid ‘binding pocket’ for isopeptag to dock in to and

react. Furthermore, I also investigated reducing the length of isopeptag by 5 residues at its N-terminal end (Fig. 3.8a). This removed the residues in isopeptag that were part of the loop region and also the first residue of the β-strand (Fig. 3.8a). Extending the length of the binding partner to include residues already present in isopeptag caused an overlapping of the duplicated residues, and I wanted to investigate the effect this overlap had on the efficiency of the reaction between isopeptag and its binding partner.

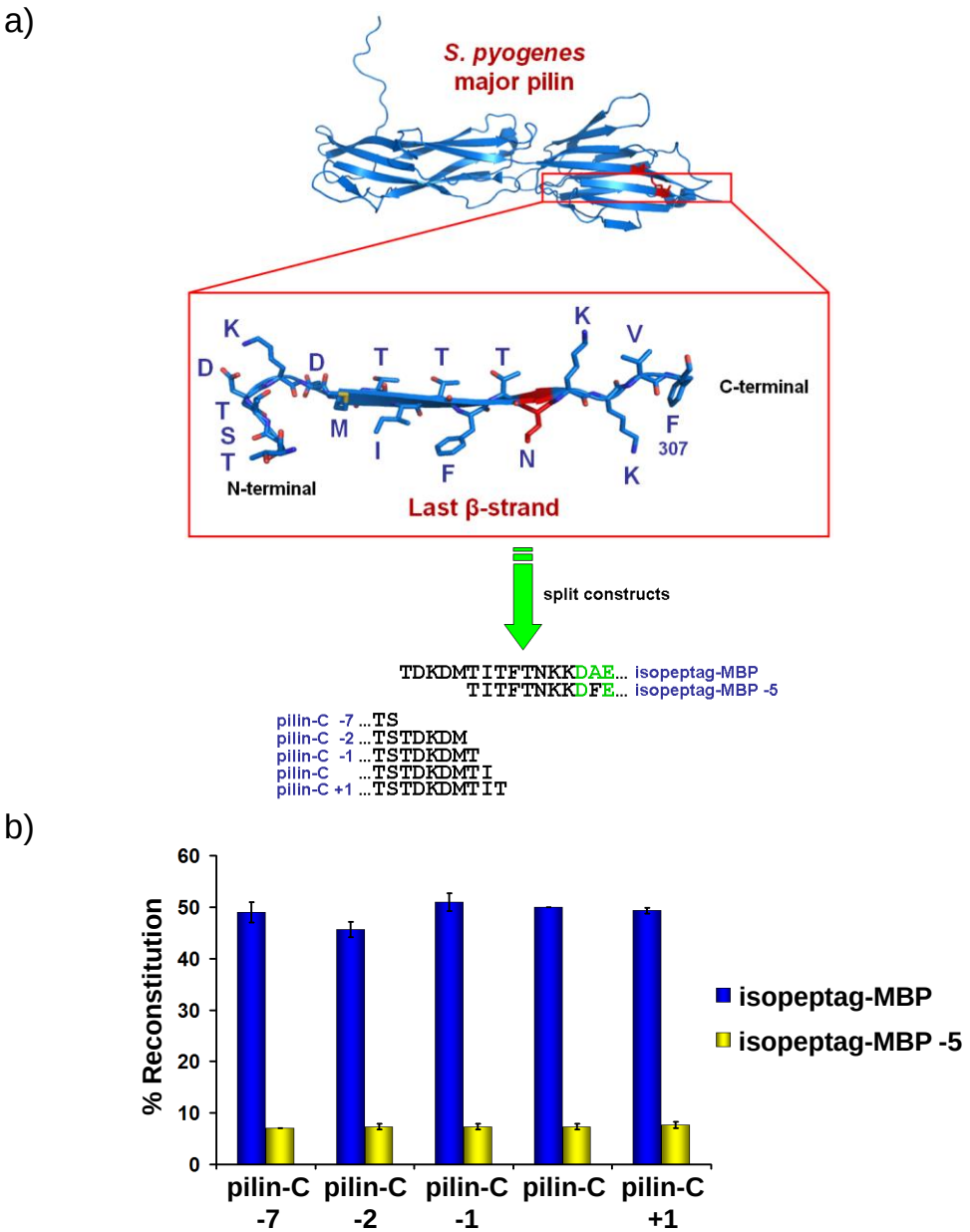


Figure 3.8. Optimization of the lengths of isopeptag and pilin-C -7. a) Schematic of the major pilin subunit Spy0128 (PDB 3B2M) from *S. pyogenes* illustrating the various isopeptag-MBP and pilin-C variants generated. Phe^{307} is the final residue visible on the

available crystal structure of Spy0128. Also, the residue at position 306 in the Spy0128 structure is Val instead of Asp as in *S. pyogenes* M1 GAS (NCBI Entrez Nucleotide accession no. AE004092). Isopeptag contained an Asp at position 306 and Phe^{307} was mutated to an Ala. Asn^{303} involved in isopeptide bond formation is shown in red and residues shown in green are different to the crystal structure of Spy0128. The numbering of the isopeptag-MBP and pilin-C -7 variants illustrate the change in the number of amino acids relative to the optimized constructs of isopeptag and pilin-C. b) To assess the reconstitution efficiency of the various isopeptag-MBP and pilin-C -7 constructs, each protein was mixed at a concentration of 10 μM in PBS pH 7.4 and incubated at 25ºC for 24 hr. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining. Reactions were performed in triplicate and the means of results are shown ± 1 s.d.
This figure was published in (Zakeri and Howarth, 2010).

3.4.4. Reconstitution efficiency of the isopeptag and pilin-C -7 variants

To investigate how altering the length of pilin-C -7 affected its reconstitution efficiency with isopeptag, I extended the C-terminal end of the protein by 5, 6, 7, and 8 amino acids (Fig. 3.8a). This incrementally increased the length of the binding partner until it extended half way in to the C-terminal β-strand (Fig. 3.8a). I also reduced the length of isopeptag at its N-terminal end by 5 amino acids (Fig. 3.8a), and then incubated the isopeptag-MBP and pilin-C -7 variants at 10 μM of each protein in PBS pH 7.4 for 24 hr at 25°C, and then analyzed the reactions on SDS-PAGE (Fig. 3.8b). Increasing the length of pilin-C -7 did not have a significant effect on the efficiency of reconstitution (Fig. 3.8b). However, the more similar the split protein is to wild-type Spy0128, the more stable a structure the split protein is likely to have. Therefore, I decided to do all further experiments with pilin-C because its sequence extended in to the final β-strand but was still four residues away from the reactive Asn (Fig. 3.8a). Also, pilin-C provided a larger binding surface area for isopeptag than pilin-C +1 (Fig. 3.8a). On the other hand, reducing the length of isopeptag by 5 residues was severely detrimental to reconstitution efficiency with the pilin-C variants (Fig. 3.8b). Although isopeptag -5 still contained all but one residue in the final β-strand (Fig. 3.8a), a significantly reduced reactivity implied that the preceding sequences in the loop region were critical to the interaction between isopeptag and the protein binding partner. Hence, isopeptag and pilin-C constituted the best interacting peptide-protein pair and were used for all further experiments.

3.5. Characterization of isopeptag and pilin-C reactivity

To gain a deeper understanding of the interaction between isopeptag and pilin-C (Fig. 3.9a), I sought to characterize the reaction between the peptide-protein pair.

3.5.1. Isopeptag-MBP and pilin-C specifically reconstituted in vitro

When isopeptag-MBP and pilin-C were each mixed at a concentration of 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25°C for 24 hr, SDS-PAGE analysis revealed a new band at ~80 kDa representing a covalently linked isopeptag-MBP:pilin-C complex (Fig. 3.9b). As a negative control, the reactive Lys^{179} in pilin-C was mutated to an Ala to make pilin-C K179A. However, when isopeptag-MBP and pilin-C K179A were incubated together, no covalent complex was observed demonstrating that covalent complex formation required Lys^{179} (Fig. 3.9b). Also, to test the specificity of pilin-C for isopeptag I incubated pilin-C with MBP-isopeptag-N (Fig. 3.4b), a peptide that contained four Asn residues that could potentially react with Lys^{179} in pilin-C. However, incubation of pilin-C with MBP-isopeptag-N did not yield any new product demonstrating that pilin-C did not react non-specifically with Asn containing peptides (Fig. 3.9b). Interestingly, a second band was present in purified pilin-C at ~75 kDa (Fig. 3.9b), which may have been a result of covalent dimerization of pilin-C.

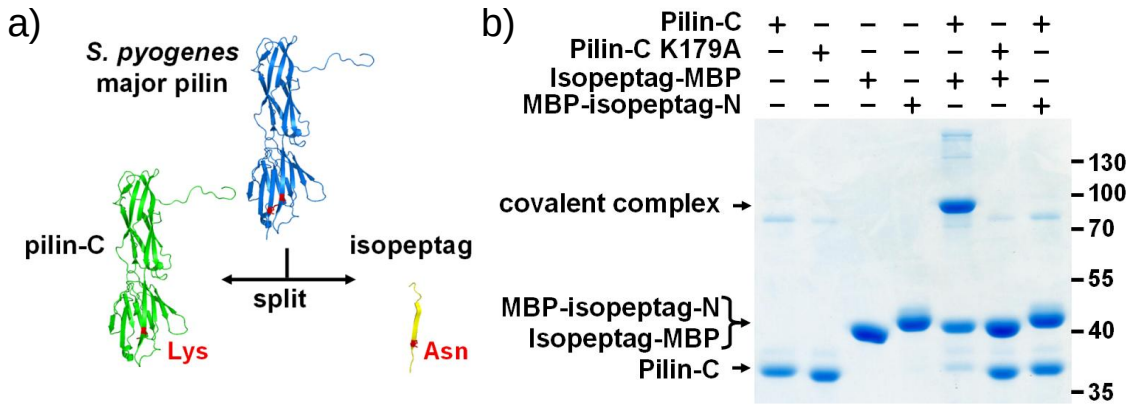


Figure 3.9. Covalent reconstitution between isopeptag-MBP and pilin-C. a) Schematic illustrating the construction of pilin-C (green) and isopeptag (yellow) by splitting the *S. pyogenes* major pilin Spy0128 (PDB 3B2M). Isopeptide bond forming residues are

shown in red. b) Isopeptag-MBP and pilin-C were able to spontaneously reconstitute and form a covalent complex. Constructs were each mixed at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25°C for 24 hr. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining. As negative controls, pilin-C K179A (with the reactive Lys^{179} mutated) and the random peptide MBP-isopeptag-N were reacted with isopeptag-MBP. A band at ~75 kDa for purified pilin-C may be a result of covalent dimerization of two pilin-C molecules. This figure was published in (Zakeri and Howarth, 2010).

3.5.2. Mass spectrometric analysis of the isopeptag-MBP and pilin-C reaction

I used mass spectrometry to demonstrate the covalent nature of the intermolecular interaction between isopeptag-MBP and pilin-C (Fig. 3.10). The expected molecular mass for isopeptag-MBP was 45,118.8 Da and for pilin-C was 33,195.8 Da, which were in close agreement with observed masses of 45,121.6 Da and 33,197.4 Da as detected by mass spectrometry (Fig. 3.10a and c). If isopeptag-MBP and pilin-C reacted through isopeptide bond formation, a NH_3 molecule (17 Da) would be released and the covalently complexed product would have an expected molecular mass of 78,297.6 Da (45,118.8 Da + 33,195.8 Da – 17 Da). Mass spectrometric analysis of the reaction between isopeptag-MBP and pilin-C yielded a new signal at 78,299.8 Da, in close agreement with the expected mass of an isopeptide linked isopeptag-MBP:pilin-C complex (Fig. 3.10d). However, when isopeptag-MBP was incubated with the unreactive pilin-C K179A construct and subsequently analyzed by mass spectrometry, a covalent complex was not observed (Fig. 3.10e). Therefore, isopeptag-MBP and pilin-C reconstituted via a spontaneously formed intermolecular isopeptide bond.

Interestingly, mass spectrometric analysis of pilin-C also revealed a weak signal at 66,394.3 Da corresponding to a covalently linked pilin-C dimer (Fig. 3.10a), which was also visible on SDS-PAGE (Fig. 3.9b). While each pilin-C molecule contains one intramolecular isopeptide bond in its N-domain, the mass of the observed pilin-C dimer corresponded to two pilin-C molecules containing two isopeptide bonds. This may have been a result of one intramolecular isopeptide bond forming in the N-domain of only one of the monomers and an intermolecular isopeptide forming between the two pilin-C monomers.

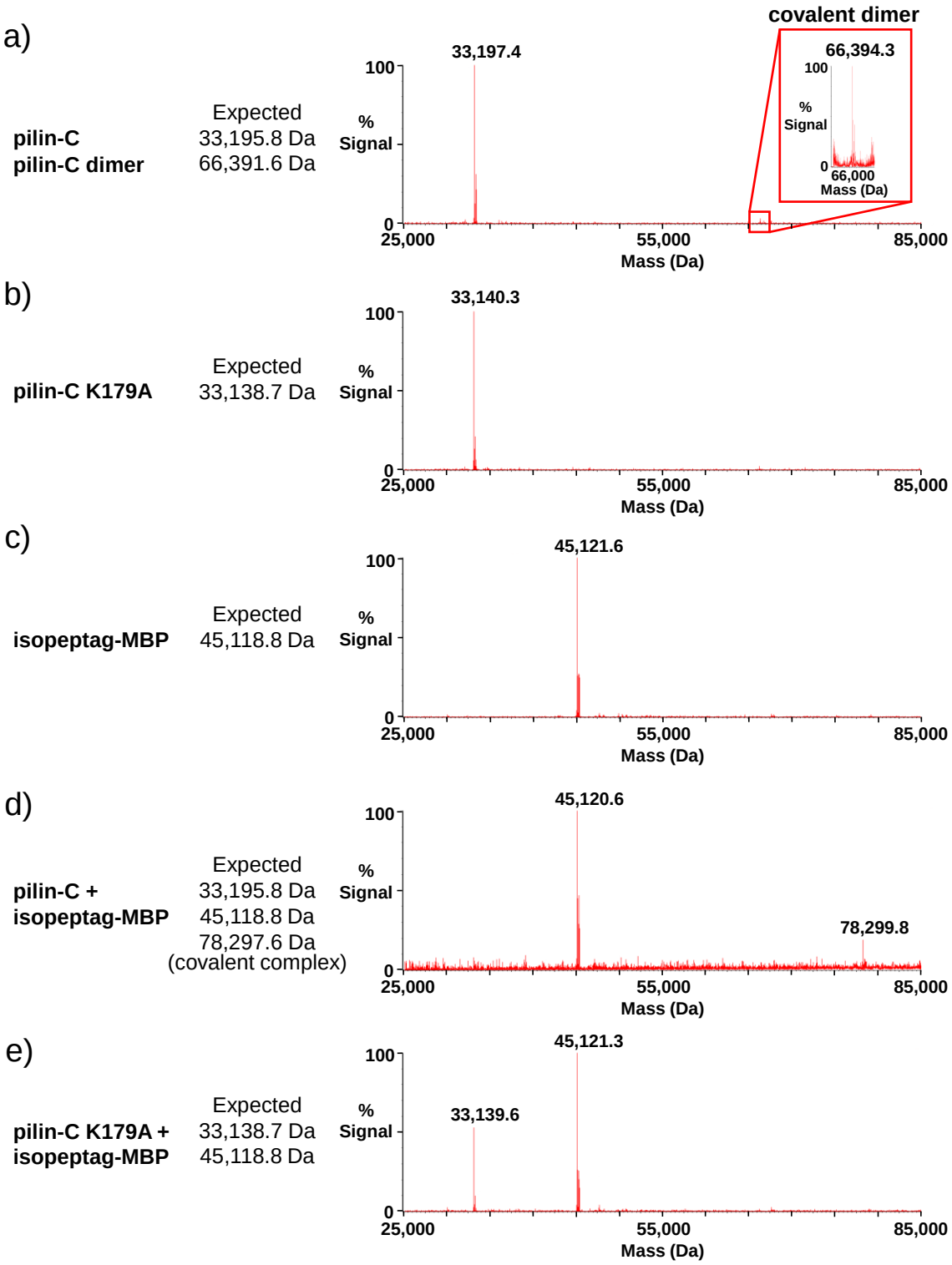


Figure 3.10. Mass spectrometric analysis of isopeptag-MBP and pilin-C reconstitution. a) Pilin-C. SDS-PAGE analysis of purified recombinant pilin-C (Fig. 3.9b) showed a band corresponding to a covalent pilin-C dimer. Mass spectrometric analysis of pilin-C also revealed a covalent dimer of pilin-C (inset). b) Pilin-C K179A. c) Isopeptag-MBP. d) Isopeptag-MBP and pilin-C reconstitution. Constructs were incubated each at a concentration of 100 μM in PBS pH 7.4 at 4ºC for 16 hr. A covalent isopeptag-MBP:pilin-C complex was detected with some unreacted isopeptag-MBP. e) Isopeptag-MBP and pilin-C K179A were reacted as in (d), however no covalent complex was observed.
This figure was published in (Zakeri and Howarth, 2010).

3.5.3. Time course of isopeptag-MBP and pilin-C reconstitution

I next proceeded to perform a time course experiment to analyze isopeptag-MBP and pilin-C reconstitution over a 72 hr period. I mixed 10 µM of each construct in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0, and incubated the reaction at 25°C. Subsequently, I removed samples at various time points and analyzed them on SDS-PAGE (Fig. 3.11). Covalently complexed product could be detected after 1 hr of reaction and the reaction proceeded steadily until 3 hr, after which it began to slow down and plateau at a maximum of ~60% reconstitution after 72 hr (Fig. 3.11a). One possible explanation for why the reaction did not proceed to completion could be that the destabilized pilin-C construct may have been present in several different conformations. Notably, purified pilin-C was always accompanied by another protein, which I believe represented a covalent homodimer (Fig. 3.10a). Furthermore, reaction between isopeptag-MBP and pilin-C resulted in the formation of two new bands above the 130 kDa marker (Fig. 3.11b). Although these side products have not been characterized, they could be the result of reaction between the pilin-C dimer and one or two isopeptag-MBP molecules. Another possibility for why the reaction did not proceed to completion could be that the isopeptag peptide at the N-terminal end of MBP could also have been present in multiple conformations. Ideally the peptide would be flexible and displayed freely in solution so that it could bind pilin-C. However, isopeptag has several hydrophobic residues which may interact with surrounding hydrophobic surfaces or fold in a way so as to render the peptide unable to recognize and bind pilin-C. It is unlikely for intramolecular isopeptide bond formation in Spy0128 to require several hours, because in the appropriate medium *S. pyogenes* has a doubling time of 32 min (van de Rijn and Kessler, 1980). Therefore, there is significant room for improvement in the efficiency of covalent reconstitution and other split scaffolds may reconstitute faster.

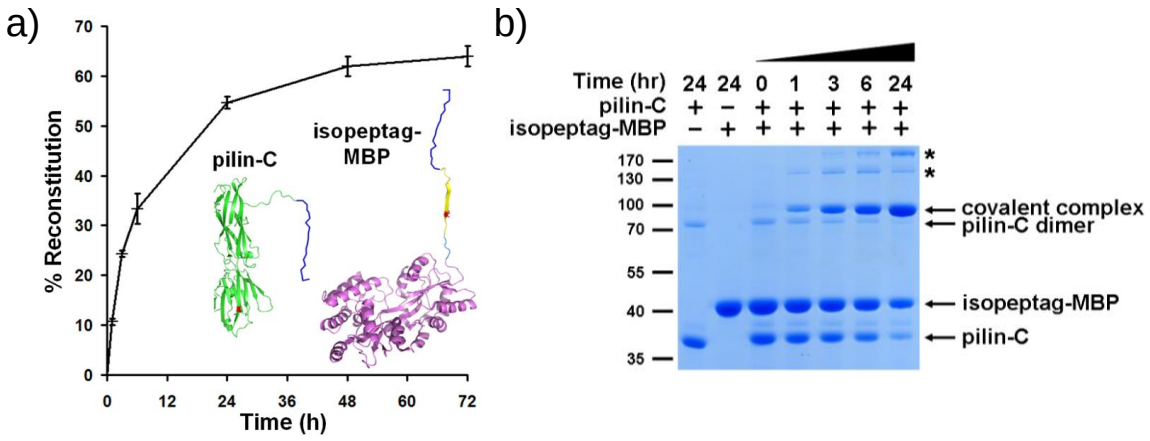


Figure 3.11. Time course of isopeptide bond formation. a) Analysis of isopeptag-MBP and pilin-C reconstitution over a 72 hr period. Isopeptag-MBP and pilin-C were incubated each at 10 μM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 at 25 °C. At various time points samples were removed and heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie Staining. Reactions were performed in triplicate, where the mean ± 1 s.d is shown. A cartoon of isopeptag-MBP and pilin-C is shown on the graph where the 6X His tags of the constructs and linkers are shown in blue. b) A representative gel of the time course experiment shown in (a).
*Side products of the isopeptag-MBP and pilin-C reaction possibly a result of 1 or 2 isopeptag-MBP molecules reacting with the pilin-C dimer.
This figure was published in (Zakeri and Howarth, 2010).

3.5.4. Isopeptag-MBP and pilin-C reconstitution at varying ratios

When isopeptag-MBP and pilin-C were mixed in equal concentrations, the reaction proceeded to a maximum product yield of ~60% (Fig. 3.11a). However, the reaction may have proceeded further if one of the reactants was present in excess. To test this, I maintained one partner at a constant concentration of 10 µM and incubated it with 0,1, 2.5, 5, 10, and 20 µM of its reacting partner. Reactions were incubated in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 at 25°C for 24 hr, and then analyzed by SDS-PAGE, where the disappearance of each reactant was measured and quantified (Fig. 3.12). Interestingly, when 2-fold excess isopeptag-MBP was present ~98% of pilin-C was able to react (Fig. 3.12). On the other hand, when pilin-C was present in 2-fold excess ~85% of isopeptag-MBP reacted (Fig. 3.12). This demonstrated that even though the isopeptag-MBP and pilin-C reaction did not proceed to completion when they were mixed in equal concentrations, possibly because they may have adopted conformations not favourable for reconstitution, the reaction was able to proceed to almost completion by providing one

reactant in excess. Therefore, almost all of the isopeptag-MBP and pilin-C in a reaction are present in a form that is active and able to reconstitute if conditions are favourable.

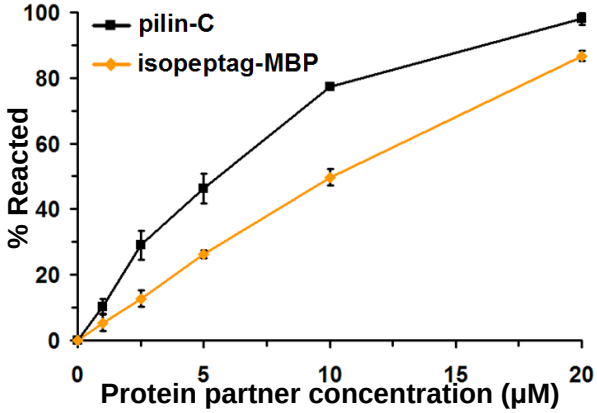


Figure 3.12. Reconstitution of one reactant against varying concentrations of another. Isopeptag-MBP (orange) or pilin-C (black) were mixed at 10 μM with varying concentration of the reacting partner. Reactions were incubated in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 at 25°C for 24 hr. Reaction were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie Staining, where the depletion of each reactant was measured. Reactions were performed in triplicate, where the mean ± 1 s.d is shown.
This figure was published in (Zakeri and Howarth, 2010).

3.5.5. Concentration-dependence of reconstitution between isopeptag-MBP and pilin-C

For many experimental conditions a protein concentration of less than 10 μM is required. Therefore, it was important to investigate the concentration-dependence of reconstitution and establish whether reconstitution could still be detected when both reactants were present at lower concentrations. Accordingly, I incubated isopeptag-MBP and pilin-C at equal concentration of 1, 5, or 10 μM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 at 25°C for varying times, and then analyzed the reactions by SDS-PAGE (Fig. 3.13). Not surprisingly, the extent of reconstitution increased with increasing concentration of the reactants (Fig. 3.13). Even though at 1 μM the reaction was least efficient, reconstitution was still detected as early as 3 hr on a SDS-PAGE gel, while the reaction proceeded to ~20% yield after 24 hr (Fig. 3.13). When isopeptag-MBP and pilin-C were present at 5 μM, reconstitution efficiency was slightly reduced as compared to

when they were present at 10 µM (Fig. 3.13). After a 24 hr reaction, ~45% covalent complex formation was observed at 5 µM, whereas ~55% was observed at 10 µM (Fig. 3.13).

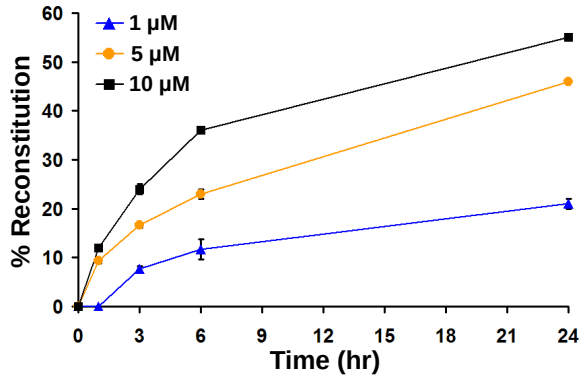


Figure 3.13. Concentration-dependence of reconstitution between isopeptag-MBP and pilin-C. Isopeptag-MBP and pilin-C were mixed at equal concentrations of 1, 5, or 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25 ºC. Samples were removed at varying time points and heated to 95°C for 7 min in SDS loading buffer, then analyzed on 10% SDS-PAGE with Coomassie Staining. Reactions were performed in triplicate, where the mean ± 1 s.d is shown.
This figure was published in (Zakeri and Howarth, 2010).

3.5.6. Temperature-dependence of isopeptag-MBP and pilin-C reconstitution

A biotechnological tool must be robust and able to function under most experimentally relevant conditions, where it should not be the limiting factor in the experiment in which it is being employed. An attractive feature of the isopeptag:pilin-C system is its versatility. Since the reaction is spontaneous, the peptide-protein pair can be used *in vitro* to stability link biological molecules in an intermolecular reaction without the requirement of any accessory factors. Often, experiments involving proteins are be conducted at 4°C to minimize protein degradation. Alternatively, when working with stable samples a molecular tool that can function at room temperature would be very convenient. Moreover, isopeptag and pilin-C can be genetically encoded allowing for *in vivo* studies to be conducted in bacterial and mammalian cells, therefore it would be

advantageous for them to be able to react at 37°C. As such, I tested the ability of isopeptag-MBP and pilin-C to react at various temperatures by mixing each construct at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubating the reactions at 4, 25, and 37°C for 3 and 24 hr. Reactions were then analyzed on SDS-PAGE (Fig. 3.14). Unexpectedly, isopeptag-MBP and pilin-C covalent reconstitution was largely temperature-independent (Fig. 3.14). Similar amounts of covalent complex formation was observed at 4°C and 25°C, while slightly lower at 37°C (Fig. 3.14). Likewise, the speed of reconstitution was only a little affected by temperature, where reconstitution efficiency at different temperatures was similar at varying time points (Fig. 3.14).

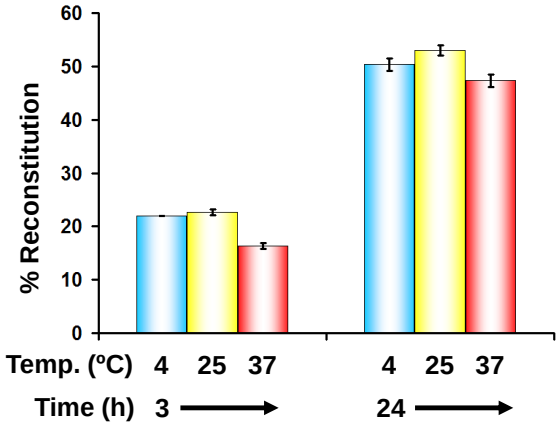


Figure 3.14. Temperature-dependence of isopeptag-MBP and pilin-C reconstitution. Isopeptag-MBP and pilin-C were each mixed at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 4, 25, and 37°C. Samples were removed at 3 and 24 hr and heated to 95°C for 7 min in SDS loading buffer, then analyzed on 10% SDS-PAGE with Coomassie Staining. Reactions were performed in triplicate, where the mean ± 1 s.d is shown.
This figure was published in (Zakeri and Howarth, 2010).

3.5.7. pH-dependence of isopeptag-MBP and pilin-C reconstitution

A peptide fusion tag should be functional in a range of pH values that are biologically relevant. Consequently, I proceeded to test whether isopeptag-MBP and pilin-C reconstitution was pH-dependent. I mixed each construct at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid with the buffer pH adjusted to 5, 6, 7, and 8. Subsequently, I incubated

the reactions at 25°C for 3 and 24 hr followed by analysis on SDS-PAGE (Fig. 3.15). Interestingly, isopeptag-MBP and pilin-C were able to covalently reconstitute at all the pH values tested (Fig. 3.15). Similar levels of reconstitution were observed at pH 7 and 8, while efficiency was slightly reduced at pH 6 (Fig. 3.15). Reconstitution was least efficient at pH 5, where after a 24 hr incubation period ~15% less product was attained as compared to pH 7 and 8 (Fig. 3.15). Furthermore, the speed of reconstitution was only a little affected by buffer pH, where after both 3 and 24 hr incubation periods pH 7 and 8 were most efficient, followed by pH 6 and then pH 5 (Fig. 3.15). Therefore, isopeptag-MBP and pilin-C reconstitution was able to proceed efficiently at a range of biologically relevant pH values.

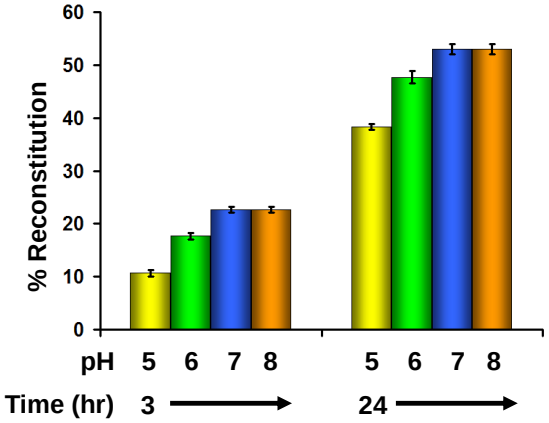


Figure 3.15. pH-dependence of isopeptag-MBP and pilin-C reconstitution. Isopeptag-MBP and pilin-C were each mixed at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid with buffer pH adjusted to pH 5.0, 6.0, 7.0, and 8.0. Reactions were then incubated 25°C and samples were removed at 3 and 24 hr, then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie Staining. Reactions were performed in triplicate, where the mean ± 1 s.d is shown.
This figure was published in (Zakeri and Howarth, 2010).

3.5.8. Isopeptag-MBP and pilin-C reconstitution in various buffers and detergents

Other important parameters for a robust biotechnological tool are its ability to function in various biological buffers and in the presence of commonly used detergents. Ideally, the choice of a biological buffer should be based on the requirements of the system under investigation and not on the peptide tag fused to the target protein. Furthermore, detergents are sometimes added to buffers to break cellular membranes or

disrupt non-specific protein interactions, and it would be advantageous to have a peptide tag that can still bind its protein partner under these conditions. Therefore, I tested the ability of isopeptag-MBP and pilin-C to covalently reconstitute in the presence of commonly used biological buffers and detergents (Fig. 3.16). Isopeptag-MBP and pilin-C were mixed at 10 µM in PBS pH 7.4, 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.4, Tris pH 7.4, or HEPES pH 7.4, where buffer pH was adjusted using either NaOH or KOH to test if Na^+ or K^+ was required for isopeptide bond formation. Furthermore, reconstitution was also assessed in each buffer containing either 1% Triton-X 100 or 0.5% NP40. All reactions were incubated at 25°C for 24 hr and analyzed on SDS-PAGE (Fig. 3.16). As expected, the choice of the biological buffer or the presence of Na^+ or K^+ did not have a substantial effect on the extent of reconstitution (Fig. 3.16). However, it was interesting to find that the presence of detergent did not have a substantial effect on the efficiency of reconstitution either (Fig. 3.16). Since pilin-C has hydrophobic regions that are exposed to solution, it would be reasonable to expect that the presence of detergent would have an effect on reconstitution efficiency.

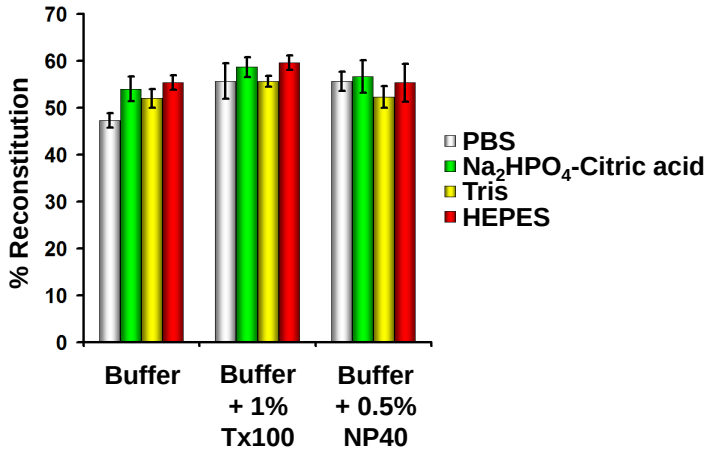


Figure 3.16. Isopeptag-MBP and pilin-C reconstitution in various buffers and detergents. Isopeptag-MBP and pilin-C were each mixed at 10 µM in PBS pH 7.4, 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.4, Tris pH 7.4, or HEPES pH 7.4 with or without 1% Triton X-100 (Tx100) or 0.5% Nonidet P-40 (NP40). Reactions were incubated at 25°C for 24 hr, then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-

PAGE with Coomassie Staining. Reactions were performed in triplicate, where the mean ± 1 s.d is shown.
This figure was published in (Zakeri and Howarth, 2010).

3.6. Isopeptag-MBP and pilin-C could reconstitute inside a bacterial cell

After characterizing the isopeptag and pilin-C system and demonstrating that the reaction was robust and could occur under most biologically relevant conditions, I next wanted to test potential applications of the technology. One application where this technology can be potentially advantageous is that it can be used as a method to permanently cross-link two proteins inside cells. Since both partners are composed of the standard 20 amino acids, they can be genetically encoded and expressed inside cells. Additionally, reconstitution is covalent, spontaneous, independent of the redox status of the environment, and does not require any accessory factors. Therefore, both isopeptag and pilin-C can be genetically linked to two proteins of interest and expressed inside a cell, and if isopeptag and pilin-C come in to contact they will be permanently linked through a covalent and irreversible bond.

Thus far, all of the experiments involving isopeptag and pilin-C were performed using purified proteins under controlled conditions. Although I had demonstrated that pilin-C was specific for isopeptag-MBP over MBP-isopeptag-N (Fig. 3.9b), I was unsure whether the reaction would still be specific within a complex environment such as a bacterial cytosol. Therefore, I decided to test whether reconstitution was possible inside the cytosol of *E. coli*.

I constructed a bicistronic vector by inserting the genes for isopeptag-MBP and pilin-C or pilin-C K179A (negative control), each with its own ribosomal binding site, in to the expression vector pET28a (Fig. 3.17a). Each vector was transformed in to the expression strain *E. coli* BL21 DE3 RIPL and protein expression was induced for 4 hr at 30°C. Samples were then taken and heated to 95°C for 7 min in SDS loading buffer and analyzed on SDS-PAGE (Fig. 3.17b). When isopeptag-MBP and pilin-C were expressed together in *E. coli*, a band at ~80 kDa representing a covalent complex between the two

constructs was observed (Fig. 3.17b). However, this band was not present when monocistronic vectors or the bicistronic isopeptag-MBP and pilin-C K179A vector were used for expression (Fig. 3.17b). Therefore, isopeptag and pilin-C can be genetically encoded and expressed inside living cells, where they are able to specifically and covalently reconstitute.

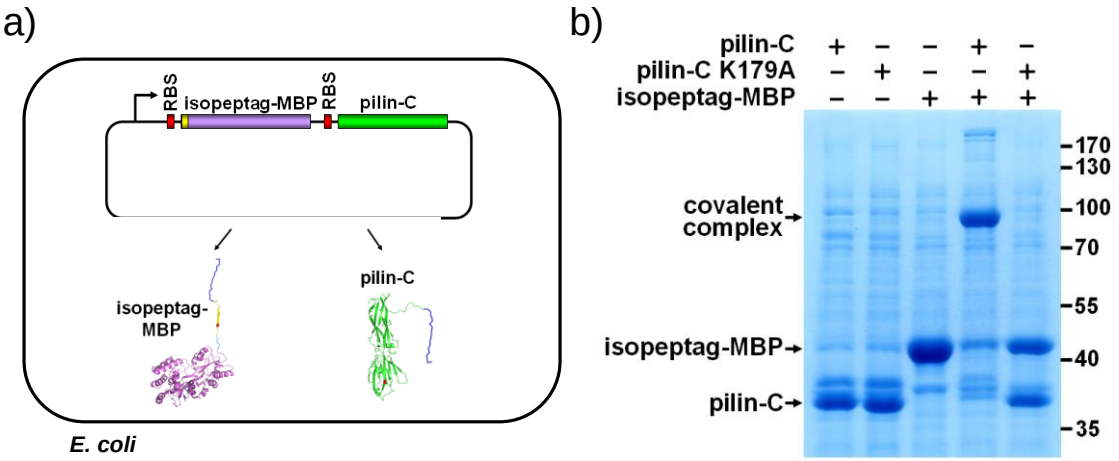


Figure 3.17. Isopeptag-MBP and pilin-C could reconstitute inside a bacterial cell. a) Cartoon illustrating the bicistronic isopeptag-MBP and pilin-C expression vector. Both constructs with separate ribosomal binding sites (RBS) were cloned in to pET28a, which was then transformed in to an *E. coli* expression strain. b) Monocistronic and bicistronic vectors were transformed in to *E. coli* BL21 DE3 RIPL and protein expression was induced for 4 hr at 30°C, after which samples were removed and heated to 95°C for 7 min in SDS loading buffer and then analyzed on 10% SDS-PAGE with Coomassie staining. Pilin-C K179A was used as a negative control, where the reactive Lys^{179} in pilin-C was mutated to an Ala to prevent isopeptide bond formation.
*Figure 3.17b was published in (Zakeri and Howarth, 2010).

3.7. Isopeptag and pilin-C were used for cell surface labelling

Fluorescent labelling of proteins for microscopic analysis is a powerful technique for detecting and monitoring the activity of various proteins in cells. This is often achieved by targeting a protein under investigation with an antibody, where the antibody is either specific for the target protein or for a peptide tag fused to the target protein. However, raising antibodies for every protein being studied is both expensive and time consuming. Therefore, it is desirable to conduct such experiments using antibodies specific for a peptide fusion tag that can be genetically linked to any protein being studied (Terpe,

2003). These antibodies are generally less expensive and can be used to detect any protein with the appropriate peptide tag.

To investigate whether isopeptag could be used for cell surface labelling experiments, I sought to demonstrate that isopeptag and pilin-C were able to reconstitute in the complex environment of a mammalian cell membrane. Accordingly, the construct isopeptag-CFP-TM was made, where isopeptag was linked to the N-terminal end of the cyan fluorescent protein (CFP) which was linked at its C-terminal end to a trans-membrane helix (TM). When this construct was expressed inside mammalian HeLa cells, the TM helix was incorporated in to the cell membrane with the isopeptag-CFP region being displayed on the plasma membrane (Fig. 3.18a). With the isopeptag being displayed on the cell surface, I added exogenous pilin-C and detected reconstitution using an antibody against the 6X His tag present at the N-terminal end of pilin-C (Fig. 3.18a). Isopeptag-CFP-TM was expressed well in HeLa cells as was apparent from the strong CFP signal (Fig. 3.18b). When exogenous pilin-C was added to isopeptag-CFP-TM expressing cells, there was clear labelling of the cells based on the strong Alexa Fluor® 555 signal (Fig. 3.18b). On the other hand, when pilin-C K179A was added to isopeptag-CFP-TM expressing cells, there was no binding detected (Fig. 3.18b). Therefore, pilin-C was able to specifically bind isopeptag in the complex environment of a mammalian cell membrane.

This experiment demonstrated the feasibility of the isopeptag:pilin-C technology for cell surface labelling experiments. The advantage of a covalent bond forming between isopeptag and pilin-C over convention antibodies, which bind through weak non-covalent interactions, allows experiments to be conducted under more stringent conditions. For example, isopeptag displaying cells could be specifically labelled using pilin-C conjugated to a fluorescent protein or dye, and then the labelled cells could be washed under harsher conditions with greater shear force to reduce background signal.

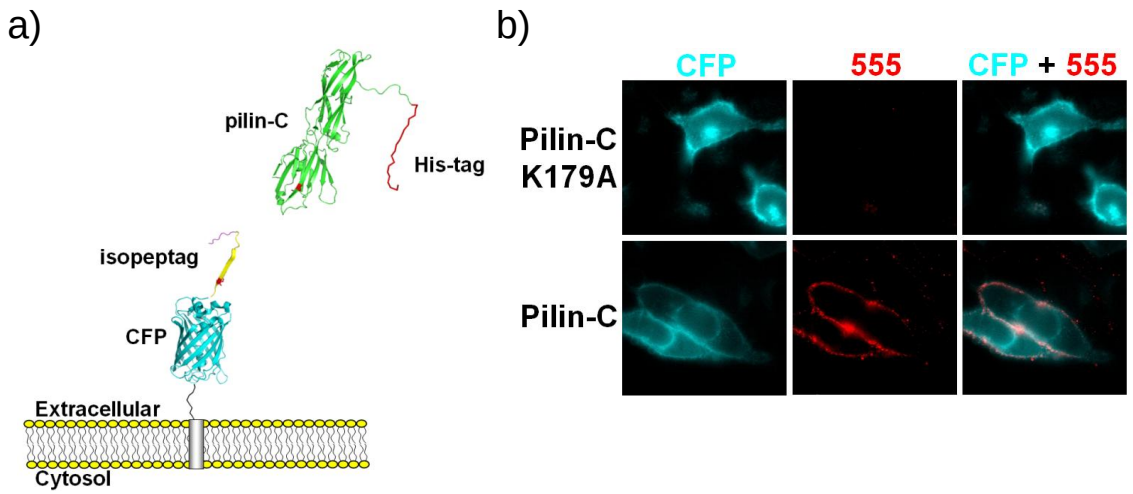


Figure 3.18. Cell surface labelling with isopeptag and pilin-C. a) A schematic illustrating isopeptag-CFP-TM incorporated in the plasma membrane of a mammalian cell. Subsequently, exogenously added pilin-C can be used to bind surface displayed isopeptag. Binding can be detected using an antibody against the 6X His tag at the N-terminal end of pilin-C. b) Isopeptag-CFP-TM (cyan) expressing HeLa cells were incubated with pilin-C or pilin-C K179A (negative control with the reactive Lys^{179} mutated to an Ala) at 5 µM for 4 hr at 37°C. Cells were then washed and reconstitution was detected using an anti-His tag primary antibody and a secondary antibody labelled with Alexa Fluor® 555 (red). In right panel, the CFP and 555 images are merged.
This figure was published in (Zakeri and Howarth, 2010).

3.8. Summary

Herein, I dissected the major pilin subunit from *S. pyogenes* to construct a covalently binding peptide fusion tag and its binding partner. I then attempted to increase the reconstitution efficiency of the peptide-protein pair by exploring alternative sizes of each partner. While, the peptide was not tolerant to a reduction in length, increasing the length of the binding partner did not affect reconstitution yield. Therefore, to make the optimized peptide protein pair I mutated a hydrophobic residue in the peptide to increase solubility to make isopeptag, and increased the size of the binding partner for increased stability to make pilin-C. I subsequently characterized the isopeptag:pilin-C reaction and demonstrated that they associated through a covalent isopeptide bond. Moreover, reconstitution was robust and occurred under all of the experimental conditions tested. Specifically, reconstitution proceeded at a range of temperatures, pH values, concentrations, buffer compositions, and in the presence of commonly used detergents. I

then proceeded to demonstrate the utility of the isopeptag:pilin-C technology by showing that they could both be genetically encoded and could reconstitute specifically inside the complex environment of a bacterial cytosol. I also showed that this system could be used for specific labelling of cell surface proteins on mammalian cells.

Interestingly, when the N-terminal domain of Spy0128 (residues 18-172) was expressed and purified as a recombinant protein, an intramolecular isopeptide bond did not form in the construct as determined by mass spectrometry (data not shown). Furthermore, a split C-terminal domain fragment of Spy0128 (residues 173-292) was unable to covalently reconstitute with CB-MBP as determined by SDS-PAGE analysis (data not shown). These results illustrate that in order for isopeptide bond formation to occur in Spy0128, both the N- and C-terminal domains must be present. It is likely that proper folding of each domain is dependent on the presence of the other domain.

Although the isopeptag:pilin-C system is novel and addresses some of the concerns of other peptide tags, it still has some major limitations. Pilin-C is a large protein with a molecular mass of ~31.0 kDa, which may cause problems if it is to be genetically linked to a protein of interest. Also, pilin-C seems to form a small amount of covalent dimer and a reaction with isopeptag leads to the formation of side-products which are undesirable. However, a more pressing concern is the efficiency of reconstitution between isopeptag and pilin-C, which reaches a maximum of ~60% yield after a 72 hr reaction. In order for this technology to have practical applications as a peptide fusion tag, a more efficiently reacting peptide-protein pair would be desirable.

To develop a more efficient covalently interacting peptide-protein pair, splitting other intramolecular isopeptide bond containing protein scaffolds could be explored. Alternatively, *in vitro* protein evolution or rational modification techniques could be used to try to improve the isopeptag:pilin-C system. Spy0128 is a large protein and since size is an important factor for many potential applications of this technology, further work on this scaffold was not ideal. Therefore, the best approach to finding a more efficiently

reconstituting peptide-protein pair was to apply the experience gained from this proof-of-principle study to split smaller intramolecular isopeptide bond containing proteins.

Chapter 4:

Investigation of smaller protein scaffolds for the development of an efficiently reconstituting and covalent peptide-protein pair

Chapter 4: Investigation of smaller protein scaffolds for the development of an efficiently reconstituting and covalent peptide-protein pair

Dissection of Spy0128 in both the N- and C-domains led to the development of the isopeptag-N:pilin-N and isopeptag:pilin-C pairs, and demonstrated the generality of my approach for the development of covalently interacting peptide-protein pairs. The difference observed in the reconstitution yields of the split N- and C-domain constructs showed that the choice of scaffold was important for efficient reconstitution. Furthermore, a protein partner for the peptide that was smaller than pilin-C would be desirable for downstream applications of this technology. Therefore, I sought to explore splitting other smaller protein scaffolds to find a more efficiently interacting peptide-protein pair.

4.1. ACE19 contained an intramolecular isopeptide bond

4.1.1. Approach for structural similarity searches for the discovery of previously unrecognized intramolecular isopeptide bond containing proteins

At the time of this investigation the only known intramolecular isopeptide bond containing protein was Spy0128. By analyzing protein crystal structures available in the Protein Data Bank (PDB), Kang et al. were also able to identify other proteins that could potentially have intramolecular isopeptide bonds that may have been unrecognized (Kang et al., 2007). Using a similar strategy to find a small protein that also contained an intramolecular isopeptide bond, I conducted a survey of the PDB using a structural template of the three residues required for isopeptide bond formation. Structural similarity searches were performed using the SPASM (SPatial Arrangements of Side chains and Main chains) software from the Uppsala Software Factory (http://xray.bmc.uu.se/usf/) (Kleywegt, 1999). The residues involved in isopeptide bond formation in Spy0128 (Lys^{36}, Asn^{168}, and Glu^{117} triad from the N-domain isopeptide bond) were used as the template for structural similarity searches (Fig. 4.1a), with the allowed substitutions of Asp for Asn and

Gln for Glu. Subsequently, the SPASM software was used to search the PDB using the template, which provided a list of hits of proteins that contained residues structurally similar to the template. As a control, it was required for the protein from which the template came to be identified as a hit. I then analyzed the results by classifying the hits based on the organism of origin and protein function. Due to the simplicity of the template there was a high rate of negative hits (containing a similar triad of residues that were not isopeptide bonded) found by the software. Taking in to consideration that intramolecular isopeptide bonds are thought to be generally present in the outer-membrane proteins of Gram-positive bacterial pathogens (Kang and Baker, 2011), I found it useful to narrow down the results using this criteria. The presence of an intramolecular isopeptide bonds in the remaining hits were then validated by examining the electron density (if it was available in the PDB) around the bond.

4.1.2. Results from structural similarity searches

Initial searches of the PDB using the isopeptide triad from Spy0128 (Fig. 4.1a) as a template did not identify any previously unrecognized intramolecular isopeptide bond containing proteins. A possible reason for this may have been that in the isopeptide template, the Lys and Asn side chains are linked (Fig. 4.1a), whereas they would not be linked in proteins where the isopeptide bond may have been unrecognized. Therefore, I decided it would be more appropriate to use a template of isopeptide bond forming residues from a protein with a previously unrecognized isopeptide bond (identified by (Kang et al., 2007)) rather than from Spy0128. Consequently, I repeated the search using a template of the isopeptide bond forming residues from a collagen binding domain from *Staphylococcus aureus* (Kang et al., 2007; Zong et al., 2005), with the allowed substitution of Asp for Asn and Glu for Asp (Fig. 4.1b). This search was successful in identifying ACE19 (PDB 2OKM) as containing a potential intramolecular isopeptide bond. Kang et al. similarly identified ACE (PDB 2Z1P), a larger protein that contains the ACE19 subdomain, as containing a potential intramolecular isopeptide bond (Kang et al., 2007).

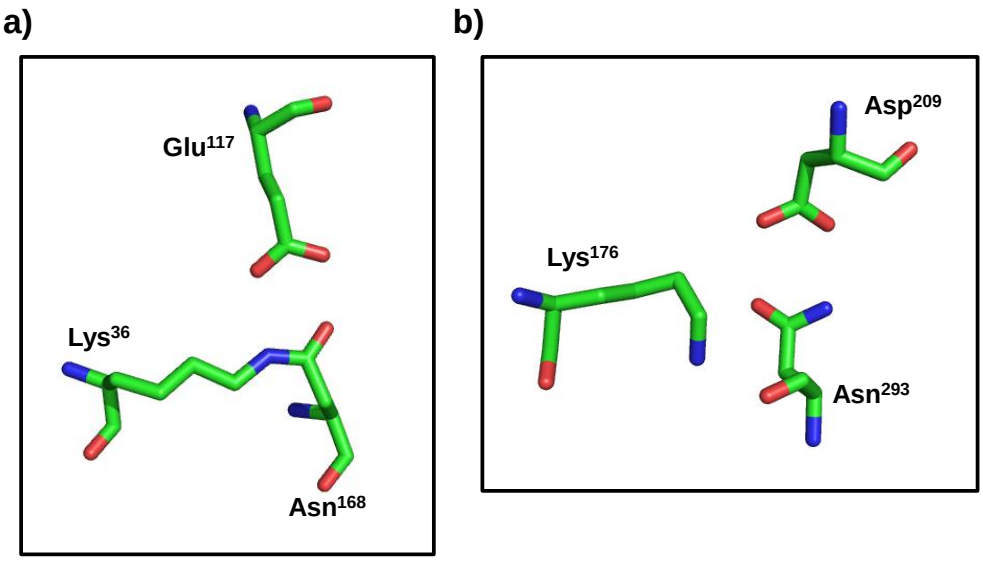


Figure 4.1. Isopeptide bond forming residues used as templates for SPASM structural similarity searches. a) Isopeptide triad from a confirmed intramolecular isopeptide bond (Kang et al., 2007). Lys[36], Asn[168], and Glu[117] in Spy0128 from *S. pyogenes* (PDB 3B2M) with allowed Asn/Asp and Glu/Gln substitutions was used as a template for structural similarity searches. b) Isopeptide triad from a previously unrecognized intramolecular isopeptide bond present in a collagen binding domain from *Staphylococcus aureus* (Kang et al., 2007; Zong et al., 2005). Lys[176], Asn[293], and Asp[209] in a collagen binding domain from *S. aureus* (PDB 2F68) with allowed Asn/Asp and Asp/Glu substitutions was used as a template for structural similarity searches. These templates were used to search the PDB with the SPASM software in order to find previously unrecognized intramolecular isopeptide bond containing proteins.

4.1.3. Further analysis of the ACE19 crystal structure

ACE19 is a 19 kDa collagen binding subdomain of the outer-membrane protein ACE (adhesin of collagen from Enterococci) from *Enterococcus faecalis* (Ponnuraj et al., 2002; Ponnuraj and Narayana, 2007; Liu et al., 2007). ACE19 is a small elongated protein with a DEv-IgG fold, and it is composed of mostly β-strands that form two β-sheets, which stack to form a β-sandwich (Fig. 4.2a) (Ponnuraj and Narayana, 2007). Closer inspections of the electron density around the side chains of the residues potentially involved in forming an isopeptide bond (Lys[42], Asn[155], and Asp[74]) as identified by the structural similarity searches (Fig. 4.2a), revealed a continuous region of electron density extending from the side chain of Lys[42] to the side chain of Asn[155] (Fig. 4.2b). More specifically, in the ACE19 crystal structure the Lys[42] ε-amino group was in a region of negative electron

density (red meshed area), whereas the space between Lys^{42} and Asn^{155} side chains was filled with positive electron density (green meshed area) (Fig 4.2b). This signified the presence of a covalent bond between the two amino acid side chains. Therefore, it was likely that a previously unrecognized isopeptide bond had formed between the side chains of Lys^{42} and Asn^{155} with the nearby Asp^{74} catalyzing the reaction (Fig. 4.2b). Since ACE19 was almost half the size of Spy0128 and contained one potential isopeptide bond (Fig. 4.3), I concluded that it was a good scaffold to further investigate.

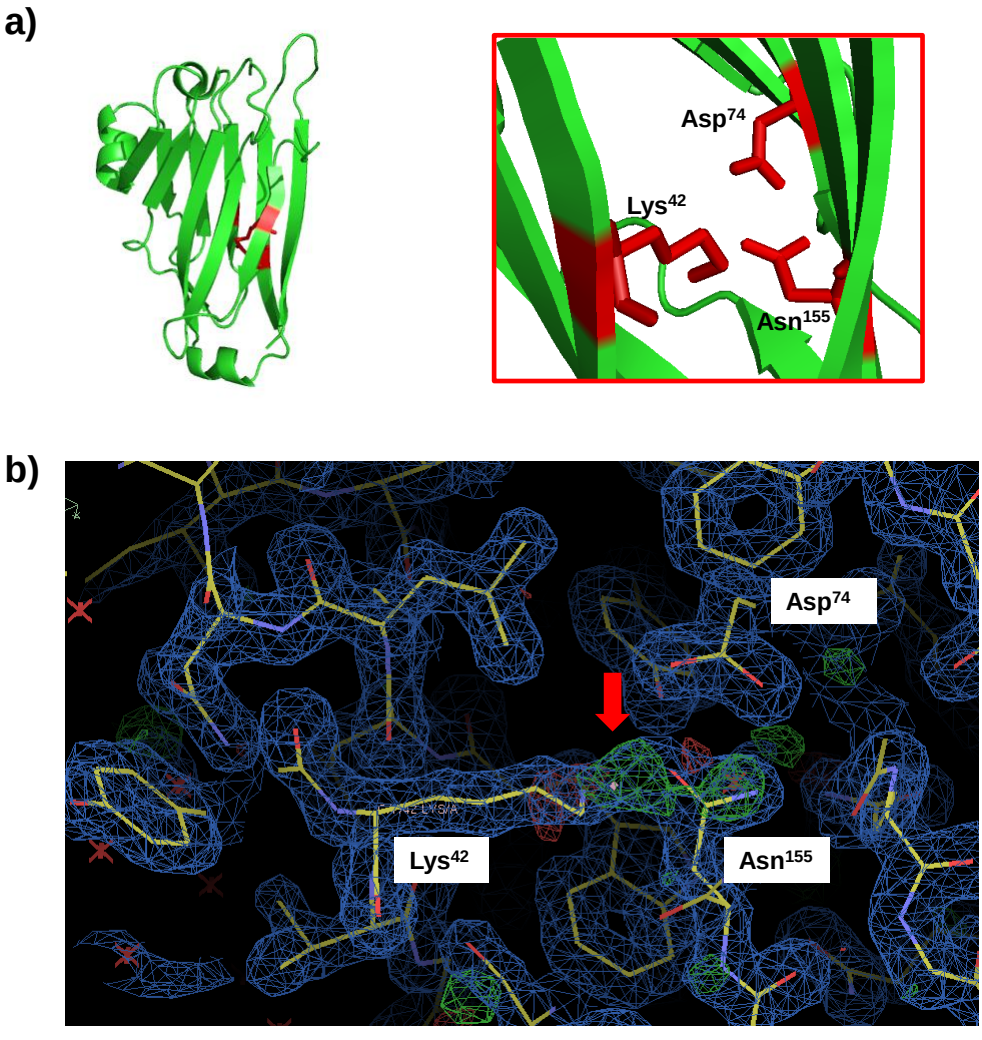


Figure 4.2. ACE19 had one potential intramolecular isopeptide bond. a) ACE19 is a 19 kDa protein composed of mostly β-strands (PDB 2OKM). It has one potential intramolecular isopeptide bond formed between Lys^{42} and Asn^{155} with the nearby Asp^{74} catalyzing the reaction (residues shown in red). A closer view of the potential isopeptide bond is shown in the red box. b) A close view of the electron density around potential isopeptide bond forming residues in ACE19. The blue mesh signifies positive electron

density around the atoms of the amino acids in ACE19. However, there was positive electron density (green mesh) between the side chains of Lys^{42} and Asn^{155} (red arrow) that was not occupied by atoms in the Lys^{42} and Asn^{155} side chains. Also, the Lys^{42} ε-amino group was positioned in a region of negative electron density (red mesh). This signified the presence of a covalent linkage between the Lys^{42} and Asn^{155} side chains. This figure was made using the software WinCoot (Emsley et al., 2010).

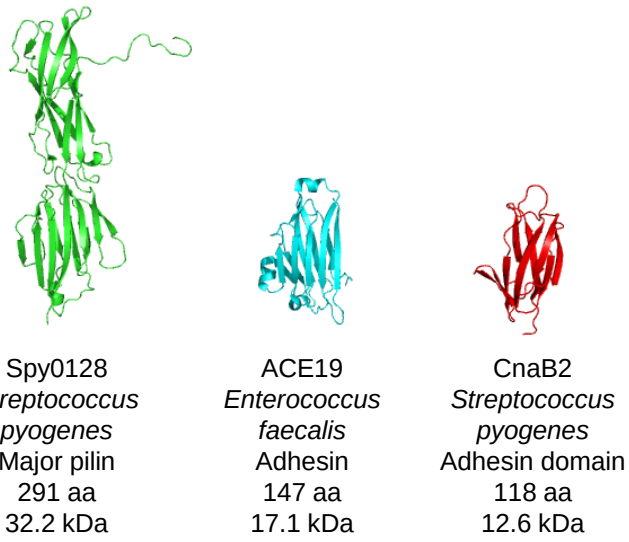


Figure 4.3. Comparison of three intramolecular isopeptide bond containing proteins. ACE19 (PDB 2OKM) and CnaB2 (PDB 2X5P) are two protein subdomains (each containing one isopeptide bond) that are smaller than Spy0128 (PDB 3B2M).

4.1.4. Mass spectrometry of ACE19

Having identified ACE19 as having a potential intramolecular isopeptide bond, I first had to verify the presence of the bond using mass spectrometry before proceeding to dissect the protein. Accordingly, I cloned the *ACE19* gene from *E. faecalis* (Andrewes and Horder 1906 strain) in to the expression vector pET28a, which fused a 6X His tag at the N-terminal end of the protein. Recombinant ACE19 expressed and purified well in *E. coli* (Fig. 4.4a). Based on its predicted sequence, recombinant ACE19 had an expected molecular mass of 19,297.9 Da. If an isopeptide bond formed between the side chains of Lys^{42} and Asn^{155} in ACE19, then NH_3 (17 Da) would be released and the subsequent protein would have an expected molecular mass of 19,280.9 Da (19,297.9 Da – 17 Da). Mass spectrometry of recombinant ACE19 revealed a protein with a mass of 19,279.8 Da

(Fig. 4.4b), in close agreement with the expected mass of ACE19 containing one intramolecular isopeptide bond. Therefore, I was successful in identifying and verifying the presence of a previously unrecognized isopeptide bond in ACE19. Subsequently, I proceeded to dissect the protein to develop a covalent peptide-protein pair.

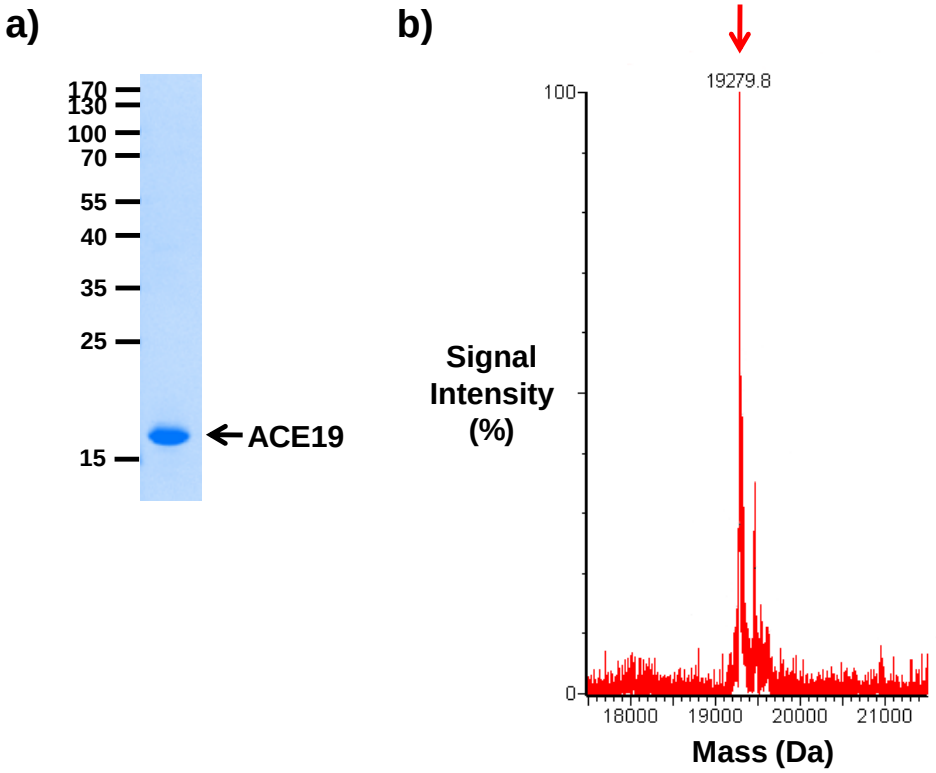


Figure 4.4. Mass spectrometry of ACE19. a) Purified ACE19 at 10 µM in PBS pH 7.4 was heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. b) Mass spectrometry of recombinant ACE19 revealed a protein with a mass of 19,279.8 Da (red arrow). Based on its predicted sequence, ACE19 had a molecular mass of 19,297.9 Da. Isopeptide bond formation in ACE19 would result in the release of a NH_3 molecule with a mass of 17 Da. Therefore, ACE19 containing an isopeptide bond was predicted to have a mass of 19,280.9 Da, in close agreement with the observed mass.

4.1.5. Dissection of ACE19

To use ACE19 as a scaffold to construct a peptide-protein pair that could covalently reconstitute via spontaneous isopeptide bond formation, I used a similar approach to the way I split Spy0128. The isopeptide bond in ACE19 linked the first and

penultimate β-strands in the protein, and thus the best approach to splitting the protein was to remove the N-terminal β-strand (Fig. 4.5a). Accordingly, I split ACE19 in the loop region after the N-terminal β-strand to generate the peptide termed ACEA, which contained the reactive Lys^{42} residue (Fig. 4.5a). ACEA was a 12 residue peptide with the sequence ERDYPFFYKVGD, constituting ACE19 residues 34-45 (sequence provided in section 2.1.3.). To aid in recombinant expression, MBP was fused to the C-terminal end of ACEA with a 3 residue linker in between. Subsequently, I made two different binding partners for ACEA, one partner named ACE44 was split inside the N-terminal β-strand and constituted ACE19 residues 44-180, and the other partner named ACE50 was split in the loop region after the N-terminal β-strand and constituted residues 50-180 in the ACE19 sequence (Fig. 4.5a). This positioned the reactive Asn^{155} and catalytic Asp^{74} residues in the two binding partners (Fig 4.5a). ACE44 overlapped in sequence with ACEA while ACE50 contained a small gap in sequence when it reconstituted with ACEA. All constructs expressed and purified well as soluble proteins in *E. coli*.

4.1.6. Split ACE19 constructs were able to covalently reconstitute

To test the ability of the split ACE19 constructs to covalently reconstitute, ACEA-MBP was mixed with either ACE44 or ACE50 at 10 μM in PBS pH 7.4 and incubated for 24 hr at 25°C. Reactions were then heated to 95°C for 7 min in SDS loading buffer to break all non-covalent interactions and analyzed by SDS-PAGE (Fig. 4.5b). As controls for covalent reconstitution, the peptide-protein pairs of PepA-MBP:Cna114 as a positive control, and PepA-MBP:Cna114 E77Q as a negative control were used (see Fig4.5b caption for details). Both ACE44 and ACE50 were able to covalently reconstitute with ACEA-MBP, although with a very low yield (Fig. 4.5b). ACE50 resulted in a slightly higher yield of a covalently complexed product than ACE44, but the yield was still far lower than the positive control of PepA-MBP and Cna114 (Fig. 4.5b), and that of the isopeptag:pilin-C pair (Fig. 3.9b). As expected, the negative control PepA-MBP and Cna114 E77Q did not result in covalent complex formation (Fig. 4.5b). Even though the split constructs

produced from the ACE19 scaffold were able to successful reconstitute via a covalent bond as determined by SDS-PAGE (Fig. 4.5b), the yield of the reaction was much lower than the isopeptag:pilin-C system. Therefore, I needed to explore splitting other small protein scaffolds to find one that could reconstitute more efficiently than the isopeptag:pilin-C pair.

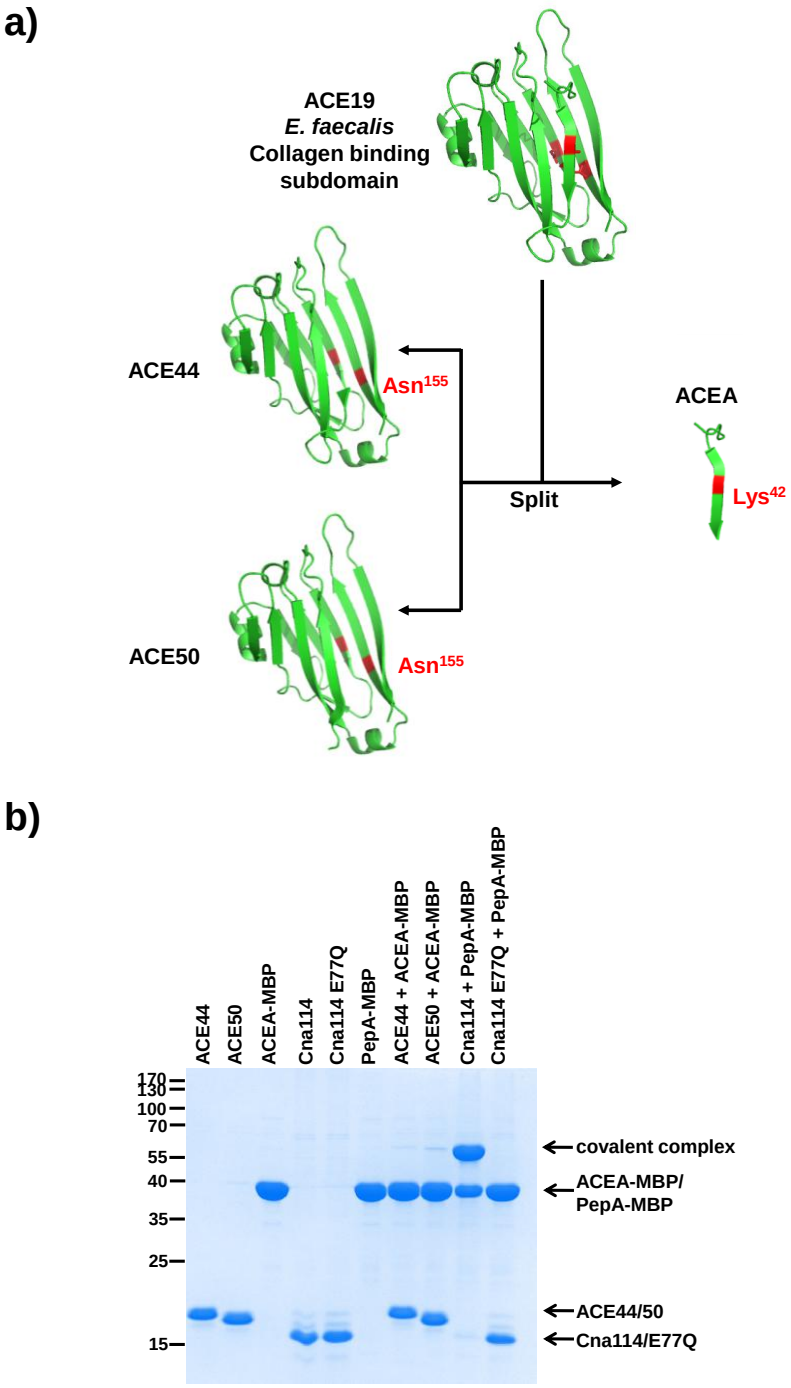


Figure 4.5. Split ACE19 constructs could covalently reconstitute. a) Cartoon illustrating the strategy used to split ACE19 (PDB 2OKM) to produce a peptide and two

binding partners. b) Covalent reconstitution of ACE19 split constructs. Proteins were mixed at 10 µM in PBS pH 7.4 and incubated for 24 hr at 25°C. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. Covalent reconstitution controls were Cna114 + PepA-MBP as a positive control and Cna114 E77Q + PepA-MBP as a negative control. These controls are based on a covalently reconstituting split peptide-protein pair developed from splitting of the intramolecular isopeptide bond containing protein CnaB2 (details about these controls are provided in sections 4.2.2. and 4.2.3.).

4.2. Investigations of the CnaB2 protein

4.2.1. CnaB2 is a small protein domain that contains an intramolecular isopeptide bond

Recently it was reported that the protein domain CnaB2 contained an intramolecular isopeptide bond (Hagan et al., 2010; Oke et al., 2010). CnaB2 is a small protein domain consisting of 118 amino acids, and is composed of all β-strands with an immunoglobulin-like fold (Fig. 4.3) (Hagan et al., 2010; Oke et al., 2010). Although the function of this particular domain is unknown, it is a part of FbaB which is a large fibronectin binding protein from *S. pyogenes* (Hagan et al., 2010). Studies of the CnaB2 protein domain revealed that it was a very stable protein that required temperatures above 100°C to denature (Hagan et al., 2010). This enhanced stability was due to the presence of an intramolecular isopeptide bond between the first and last β-strands in the protein (Fig. 4.6) (Hagan et al., 2010). Removal of this bond through mutagenesis reduced its thermal stability, where it denatured at temperatures below 60°C (Hagan et al., 2010). Interestingly, the isopeptide bond in CnaB2 formed between the side chains of Lys^{31} and Asp^{117} with a nearby Glu^{77} to catalyze the reaction (Hagan et al., 2010; Oke et al., 2010). The numbering of the residues used throughout this thesis is based on the crystal structure PDB 2X5P (Oke et al., 2010). In the published NMR structure, the residue numbering was shifted to reflect the complete FbaB protein rather than the domain itself (Hagan et al., 2010). The sequence and numbering of the residues of CnaB2 is also provided in section 2.1.3. CnaB2 represented the first example of a spontaneous intramolecular isopeptide bond formed between Lys and Asp side chains, with the leaving group being a water molecule (Hagan et al., 2010). Since CnaB2 was

almost one-third the size of Spy0128, it was an attractive candidate to further investigate for the development of a covalent peptide-protein pair.

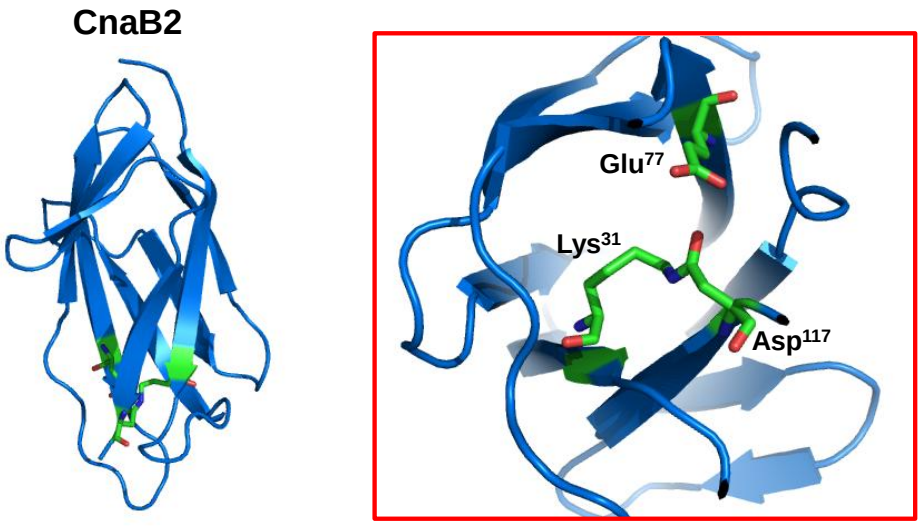


Figure 4.6. CnaB2 contains an intramolecular isopeptide bond. The N- and C-terminal β-strands in CnaB2 are linked by an isopeptide bond (residues involved in the reaction are shown in green). On the right is a closer view of the isopeptide bond formed between Lys^{31} and Asp^{117} side chains, with the nearby catalytic Glu^{77}.

4.2.2. Dissection of CnaB2

In order to develop a covalently interacting peptide-protein pair using the CnaB2 scaffold, I used a similar strategy to the way I split the Spy0128 and ACE19 proteins. The isopeptide bond in CnaB2 formed between the N- and C-terminal β-strands (Fig. 4.6). Considering the small size of the protein, my objectives were to split CnaB2 in a way that caused the minimum amount of structural perturbation. The N-terminal β-strand was composed of 8 residues, while the C-terminal β-strand contained 6 residues. Therefore, it was concluded that removal of the C-terminal β-strand may be more favourable than the N-terminal β-strand. Splitting of any protein is likely going to cause it to adopt new conformations or remain in a partially unfolded state. However, maximizing the wild-type protein sequence that is present in the split protein will likely result in the split protein folding in to a conformation resembling the wild-type protein.

Accordingly, I split CnaB2 in the loop region adjacent to the C-terminal β-strand to make the 8 residue peptide PepA, with the sequence AHIVMVDA (Fig. 4.7). PepA contained the reactive Asp^{117} residue (Fig. 4.7). To aid in recombinant protein expression, MBP was linked to the C-terminal end of PepA via a 6 residue linker. PepA had 4 hydrophobic residues in its centre, which was undesirable. These hydrophobic residues may interact with other hydrophobic surfaces either on MBP or the surrounding environment and reduce the free display of PepA in solution, thereby putting the peptide in a conformation that is less favourable for reconstitution. To address this, the two Val residues in PepA were mutated to Thr residues to make the more hydrophilic peptide PepB, with the sequence AHITMTDA (Fig. 4.7). MBP was also linked to the C-terminal end of PepB via a 6 residue linker to aid in recombinant protein expression. To make the corresponding protein partner, I split CnaB2 in the middle of the loop region prior to the final β-strand by mutating residue Asp^{110} to a stop codon to make Cna110 (Fig. 4.7). Cna110 contained the isopeptide bond forming residue Lys^{31} and the catalytic Glu^{77} residue. Similarly, I made a second protein partner, Cna114, by mutating residue Val^{114} to a stop codon (Fig. 4.7). Cna114 was split 4 residues prior to the reactive Asp^{117} inside the C-terminal β-strand.

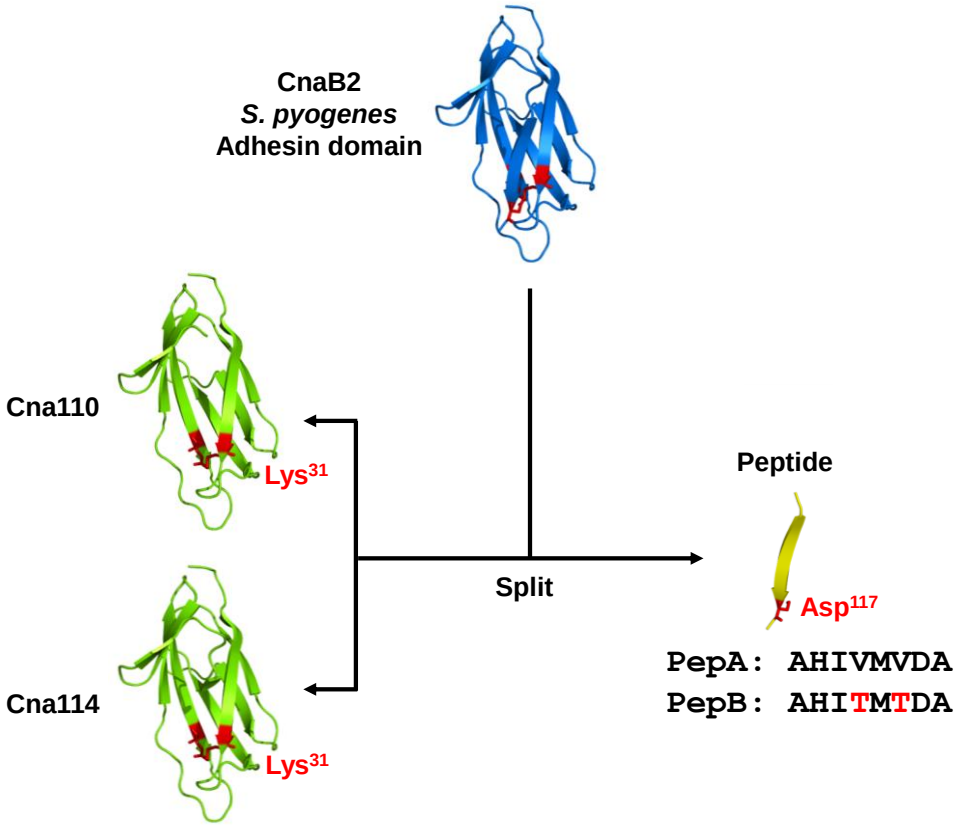


Figure 4.7. Splitting of CnaB2. CnaB2 was split to make covalently interacting peptide-protein pairs. Residues Asp^{110} and Val^{114} were changed to stop codons to make Cna110 and Cna114. Also, the C-terminal β-strand was removed to make PepA, and the two Val residues were changed to Thr residues to make PepB, a more hydrophilic peptide.

4.2.3. Split CnaB2 constructs were able to covalently reconstitute

All of the constructs expressed well in *E. coli* and purified as soluble proteins. I then mixed the peptides and binding partners at 10 μM of each in PBS pH 7.4 and incubated them for 24 hr at 25°C, and then analyzed the reactions on SDS-PAGE (Fig. 4.8a). Both PepA-MBP and PepB-MBP were able to form covalent complexes with both Cna110 and Cna114 (Fig. 4.8a). However, PepA-MBP was significantly more reactive with both binding partners than PepB-MBP, while Cna110 was slightly more reactive than Cna114 with both peptides (Fig. 4.8b). The PepA-MBP and Cna110 reaction gave a final product yield of ~83% after 24 hr, while the PepB-MBP and Cna110 reaction gave a yield of ~33% after 24 hr (Fig. 4.8b). On the other hand, the PepA-MBP and Cna114 reaction gave a product yield of ~66% after 24 hr, while the PepB-MBP and Cna114 reaction gave

a yield of ~22% after 24 hr (Fig. 4.8b). As a negative control, the catalytic Glu^{77} residue was mutated to Gln in Cna114 to make Cna114 E77Q. When Cna114 E77Q was mixed with either PepA-MBP or PepB-MBP, there was no covalent product detected (Fig. 4.8a). These results demonstrated that although the presence of the two hydrophobic Val residues in PepA may in theory have been unfavourable for solubility reasons, they were required for efficient reconstitution and isopeptide bond formation. Interestingly, in contrast to the indifference in reactivity of pilin-C to minor changes in its length (section 3.4.4.), the smaller Cna110 binding partner was more reactive than Cna114 (Fig. 4.8b). This was also seen in the split ACE19 constructs where the smaller ACE50 binding partner was slightly more reactive than ACE44 (Fig. 4.5b). In conclusion, the CnaB2 scaffold was the smallest protein that I split, and the resulting peptide-protein pair generated was the most efficient reconstituting pair tested. Thus, all further experiments were performed using the CnaB2 scaffold.

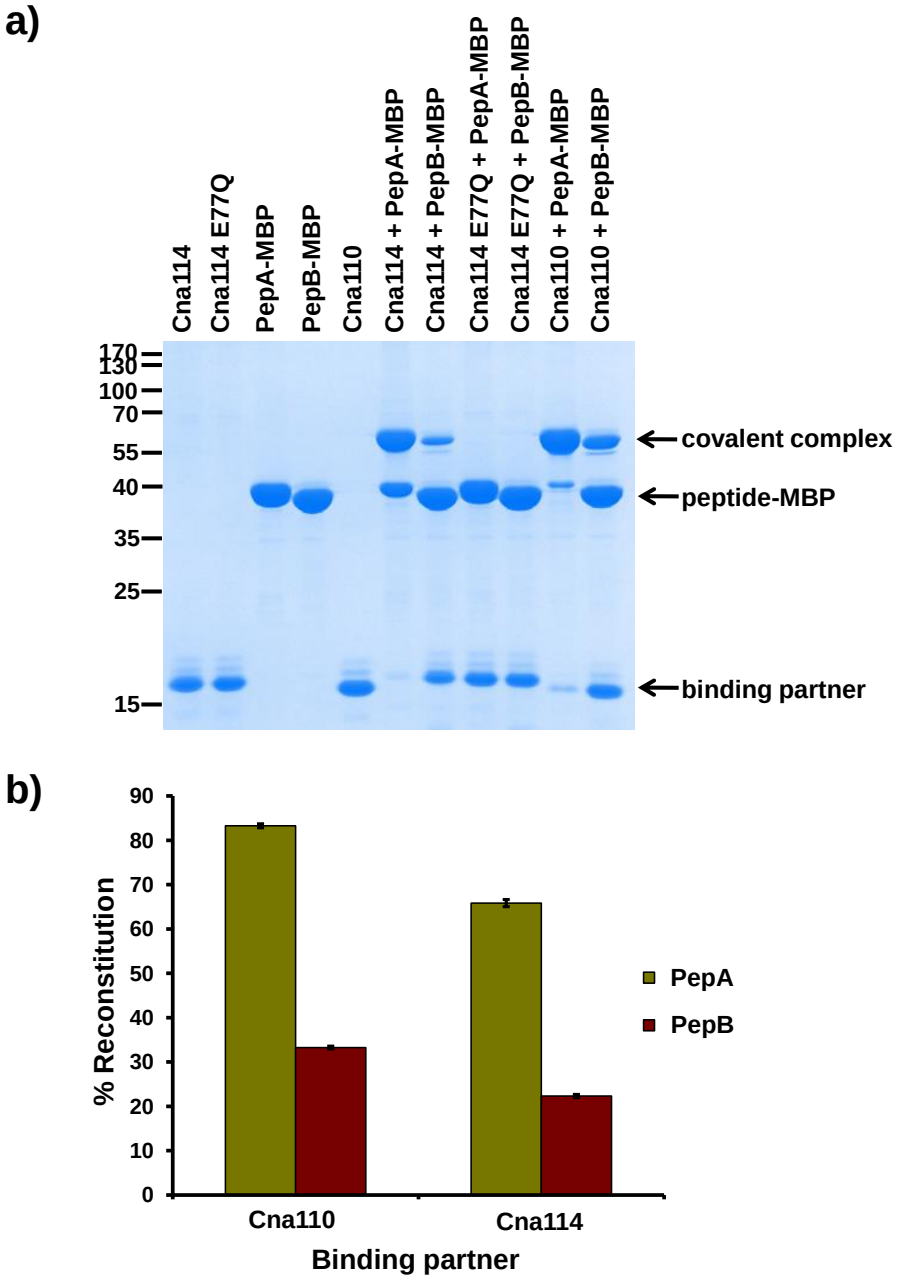


Figure 4.8. Reconstitution of split CnaB2 constructs. a) The binding partners Cna110 and Cna114 were able to covalently reconstitute with the peptides PepA-MBP and PepB-MBP. Each protein was mixed at 10 µM in PBS pH 7.4 and incubated for 24 hr at 25°C. Samples were then heated to 95°C for 7 min in SDS loading buffer and run on 14% SDS-PAGE with Coomassie staining. b) Quantification of reconstitution between Cna110 and PepA-MBP/PepB-MBP, and Cna114 and PepA-MBP/PepB-MBP. Reactions were performed in triplicate as in Fig. 4.8a and the mean is shown ± 1 s.d.

4.2.4. Preliminary characterization of the split CnaB2 reconstituting pair

Based on the reconstitution experiments conducted (Fig. 4.8), PepA and Cna110 were the most efficient covalently reconstituting peptide-protein pair. As such, I decided to perform some preliminary characterization of the PepA:Cna110 pair to gain a better understanding of their interaction.

To assess the speed of reconstitution between PepA and Cna110, I performed a time course experiment. I mixed PepA-MBP and Cna110 at 10 μM of each in 20 mM Na_2HPO_4 pH 7.4 and incubated the reaction at 25°C. At various time points samples were removed and analyzed on SDS-PAGE (Fig. 4.9a). To quantify the gel, the depletion of the Cna110 band intensity was measured (Fig. 4.9b). Cna110 depletion was fastest at earlier time points of 0-3 hr and began to slow down thereafter, and Cna110 could no longer be detected at 24 hr (Fig. 4.9b). It is interesting that at the 0 hr time point ~24% of Cna110 had reacted (Fig. 4.9b). To prepare the reaction mixture I added each protein to buffer, then mixed the reaction immediately and removed a sample for the 0 hr time point. The sample was then immediately added to SDS loading buffer and heated to 95°C for 7 min to break all non-covalent interactions. The amount of time required to mix the solution and remove the sample was ~15-20 sec, and it was fascinating that in these first few seconds about one-quarter of Cna110 had reacted.

Next, the temperature-dependence of reconstitution was assessed. I mixed PepA-MBP and Cna110 at 10 μM of each in 20 mM Na_2HPO_4 pH 7.4 and incubated the reactions at 4, 25, and 37°C for 1 hr, and then analyzed the reactions on SDS-PAGE (Fig. 4.10). Similar to the isopeptag:pilin-C pair (section 3.5.6.), reconstitution between PepA-MBP and Cna110 was not highly temperature dependent based on SDS-PAGE results, where similar product yields were obtained at 4, 25, and 37°C (Fig. 4.10).

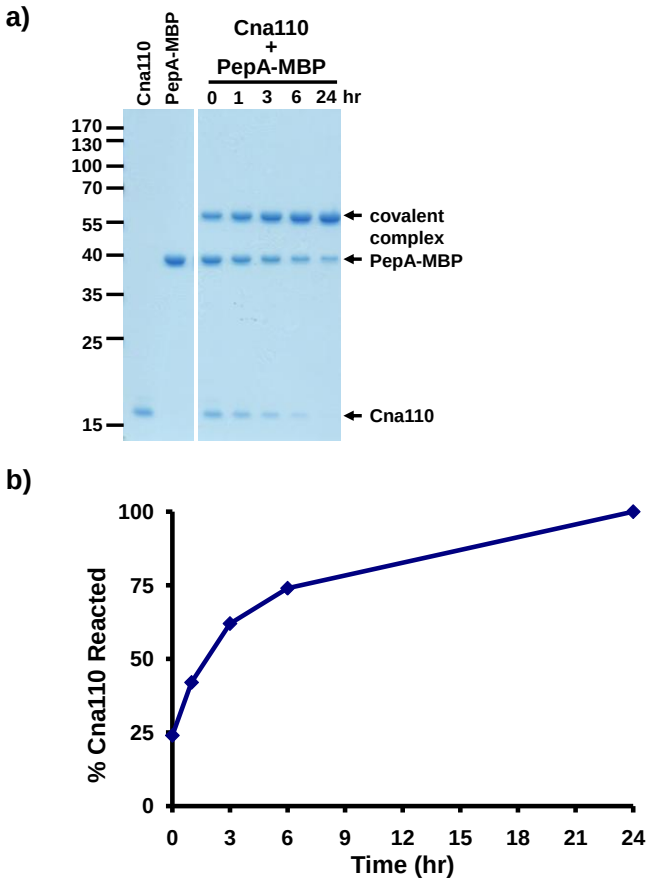


Figure 4.9. Time course of PepA-MBP and Cna110 covalent reconstitution. a) PepA-MBP and Cna110 were mixed at 10 µM in 20 mM Na_2HPO_4 pH 7.4 and incubated at 25°C. At various time points samples were removed and heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining. A lane that was not part of this experiment was removed from the gel and is shown as a white bar. b) Quantification of the PepA-MBP and Cna110 reaction shown in Fig. 4.9a.

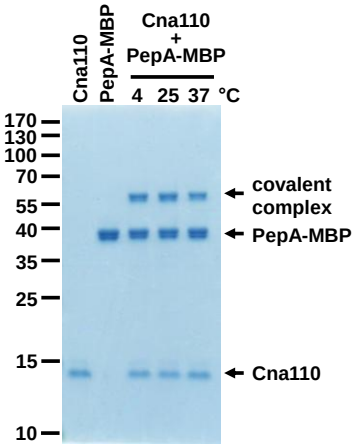


Figure 4.10. Temperature-dependence of PepA-MBP and Cna110 reconstitution. PepA-MBP and Cna110 were mixed at 10 µM in 20 mM Na_2HPO_4 pH 7.4 and incubated at 4, 25, and 37°C for 1 hr. Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining.

The split CnaB2 constructs were both smaller and more efficient than the split Spy0128 and ACE19 constructs. However, for practical applications more efficient peptide-protein pairs than the PepA:Cna110 pair would be desirable. These quick characterization experiments were performed to obtain an idea of the efficiency and the temperature-dependence of the reaction between PepA and Cna110 for downstream optimization experiments. Furthermore, an understanding of the structural aspects of reconstitution between PepA and Cna110 was valuable for optimization experiments. Therefore, we proceeded to analyze the reconstitution of split CnaB2 constructs using NMR spectroscopy.

4.2.5. Preliminary structural studies on split protein reconstitution

4.2.5.1. Cna110 and Cna peptide covalently reconstituted

To develop the covalently binding PepA:Cna110 pair, CnaB2 was split in to two fragments. By removing a β-strand from CnaB2, core hydrophobic regions of Cna110 were exposed to solution. It was unlikely that the fragmented protein would retain its wild-type structure, with a simple gap present where the β-strand was removed. The more likely scenario was that the fragmented protein would adopt a new conformation in order to reduce the exposure of its hydrophobic core to solution, and that this new conformation would be unstable. However, when the fragmented protein came in to contact with its removed β-strand in solution, the peptide-protein pair would probably interact and fold in to a conformation resembling wild-type CnaB2. I based this hypothesis on the fact that PepA and Cna110 were able to covalently reconstitute via intermolecular isopeptide bond formation, and in order for a spontaneous isopeptide bond to form, the reactive triad (Lys^{31}, Glu^{77}, and Asp^{117}) had to be positioned in exactly the correct orientation (Kang and Baker, 2011). This could only be achieved if the peptide-protein pair was able to fold to a structure resembling native CnaB2. To shed light on this process, we conducted some preliminary structural studies using nuclear magnetic resonance (NMR) spectroscopy.

To perform the NMR experiments, a commercially synthesized version of PepA was purchased with the sequence RSGAHIVMVDAGSR, termed Cna peptide (Fig. 4.11a). The PepA sequence contained four hydrophobic residues at its centre, and a concern was that PepA may not have been soluble in solution without fusion to MBP. Therefore, to make Cna peptide the hydrophilic residues RSG and GSR were introduced at N- and C-terminal ends of the peptide. Accordingly, I tested whether Cna peptide was able to form a covalent complex with Cna110. Since it was difficult to accurately determine the concentration of Cna peptide after re-suspension of the lyophilized powder, I used an excess of Cna peptide in all reactions to ensure there was enough peptide present to react with the available binding partner. To test reconstitution, I mixed 10 µM of Cna110 with 40 µM of Cna peptide in 10 mM Na_2HPO_4 pH 6.0 and incubated the reaction at 25°C for various time points, which I then analyzed using SDS-PAGE (Fig. 4.11b). I performed the reaction in 10 mM Na_2HPO_4 pH 6.0 because this was the buffer that was also used in the NMR experiments. Surprisingly, not only was Cna peptide able to react with Cna110 but the reaction proceeded rapidly, where all of Cna110 had reacted after 30 min (Fig. 4.11b). This was either because Cna peptide was present in excess or the free peptide was more reactive than peptide attached to MBP.

4.2.5.2. NMR spectroscopy of ^{15}N-Cna110 and Cna peptide reconstitution

For NMR spectroscopic analysis of reconstitution, I prepared ^{15}N-Cna110 by growing expression cultures in selective minimal media containing $^{15}NH_4Cl$ as the only source of nitrogen. To demonstrate that the activity of ^{15}N-Cna110 was not affected compared to Cna110, PepA-MBP was mixed with either Cna110 or ^{15}N-Cna110 at 10 µM in 20 mM Na_2HPO_4 pH 7.4 and the reactions were incubated at 25°C for 3 hr. Subsequently, the reactions were analyzed on SDS-PAGE (Fig. 4.11c). As expected, there was no noticeable difference between the reactivity of regular and ^{15}N-Cna110 (Fig. 4.11c). I then sent ^{15}N-Cna110 and Cna peptide to our collaborator, Dr. Uli Schwarz-Linek, for NMR spectroscopic analysis.

Detailed analysis of the NMR spectroscopic data is beyond the scope of this project and therefore will only be briefly discussed. As we have only conducted one NMR experiment on these samples, we must be cautious about drawing specific conclusions from this data. However, there are some interesting patterns that emerged from the data. NMR spectroscopic analysis of ^{15}N-Cna110 was performed using ^{1}H,^{15}N-heteronuclear single quantum coherence (HSQC) NMR. In ^{1}H,^{15}N-HSQC spectra, a cross-peak is produced for every ^{1}H-^{15}N covalent bond present in the protein, and thus there is a signal for every amino acid present in the protein backbone (due to amide bond linkages) except for prolines (Dr. Uli Schwarz-Linek, personal communication) (Ozawa et al., 2006; Schwarz-Linek, 2011). The ^{1}H,^{15}N-HSQC spectrum of wild-type ^{15}N-CnaB2 (Fig. 4.12a Box A) compared to the ^{15}N-Cna110 spectrum (Fig. 4.12a Box B) demonstrated that the two proteins were present in two different structural states. Based on the obtained spectra it cannot be determined how much of the ^{15}N-CnaB2 structure is preserved in ^{15}N-Cna110 (Dr. Uli Schwarz-Linek, personal communication) (Schwarz-Linek, 2011). However, it was clear that ^{15}N-Cna110 lacked a defined folded structure due to the large shift in the location of the W57 indole H-N (inside the hydrophobic core of the protein) signal compared to wild-type ^{15}N-CnaB2, which suggested that the hydrophobic core of ^{15}N-Cna110 was disturbed (compare Fig. 4.12a Box A and Box B) (Dr. Uli Schwarz-Linek, personal communication) (Schwarz-Linek, 2011). To test the effect of Cna peptide on the structure of ^{15}N-Cna110, 0.3 mM Cna peptide was mixed with 0.18 mM ^{15}N-Cna110 and the reaction was immediately analyzed by ^{1}H,^{15}N-HSQC NMR. There was an approximately 18 min delay from sample preparation to data acquisition. Interestingly, the ^{1}H,^{15}N-HSQC spectrum for the ^{15}N-Cna110 and Cna peptide reaction (Fig. 4.12a Box C) was more similar to the spectrum of wild-type ^{15}N-CnaB2 (Fig. 4.12a Box A) than ^{15}N-Cna110 (Fig. 4.12a Box B). This suggested that upon mixing, ^{15}N-Cna110 and Cna peptide interacted and folded in to a structure similar to the wild-type protein. An overlay of the spectra for wild-type ^{15}N-CnaB2 and ^{15}N-Cna110 + Cna peptide (Fig. 4.12a Box D) illustrated that in the reconstituted protein, the W57 signal was shifted to a location similar

to wild-type ^{15}N-CnaB2. This suggested that the hydrophobic core of the protein was restored (Dr. Uli Schwarz-Linek, personal communication) (Schwarz-Linek, 2011).

Unfortunately a signal for the isopeptide H-N bond was not present in the ^{15}N-Cna110 + Cna peptide spectrum to demonstrate the covalent nature of the interaction (Dr. Uli Schwarz-Linek, personal communication) (Schwarz-Linek, 2011). To confirm that ^{15}N-Cna110 and Cna peptide interacted covalently, our collaborators analyzed the reaction using mass spectrometry which revealed a product with a mass of 13,493 Da (Fig. 4.12b). The predicted mass for ^{15}N-Cna110 was 12,054 Da, and for Cna peptide it was 1,456 Da. If the peptide and protein reconstituted and formed an isopeptide bond then a water molecule would be released and the final product would have a predicted mass of 13,492 Da (12,054 Da + 1,456 Da – 18 Da = 13,492 Da), in close agreement with the observed mass.

Our preliminary NMR spectroscopic studies on the nature of the interaction between ^{15}N-Cna110 and Cna peptide gave us some insight in to the structural characteristics of split protein reconstitution. Although ^{15}N-Cna110 lacked a defined folded structure, in the presence of Cna peptide ^{15}N-Cna110 folded in to a structure resembling wild-type ^{15}N-CnaB2. Based on the limited data available, we could predict that the reconstituted protein was able to restore the hydrophobic core of the protein and position the catalytic residue in an orientation that was amenable for isopeptide bond formation.

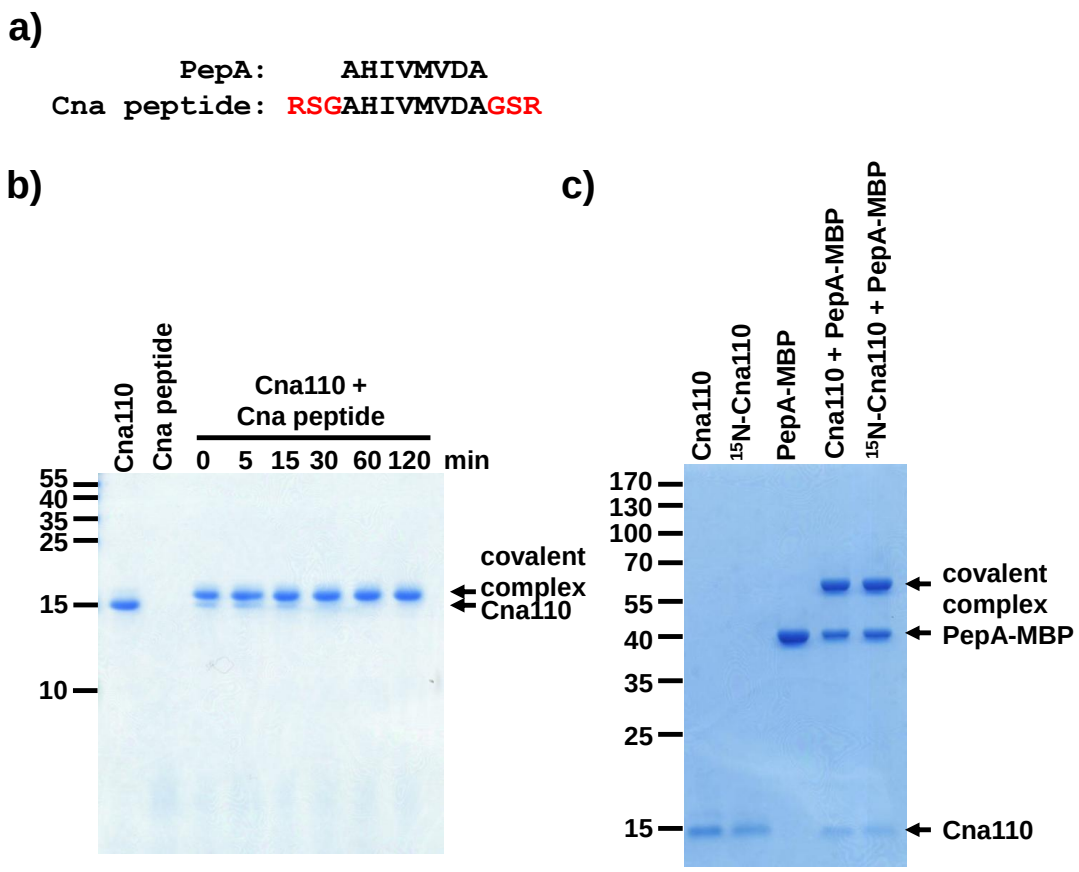


Figure 4.11. Preparation of samples for structural studies on split protein reconstitution. a) Sequence of PepA and Cna peptide. To aid in the solubility of the free peptide, RSG and GSR were added to the PepA sequence at its N- and C-terminal ends to create Cna peptide. b) Cna peptide was able to covalently reconstitute with Cna110. Cna110 at 10 µM and Cna peptide at 40 µM were mixed in 10 mM Na_2HPO_4 pH 6.0 at 25°C for various time points. Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 18% SDS-PAGE with Coomassie staining. c) Cna110 and ^{15}N-Cna110 were able to react to an equal extent with PepA-MBP. Proteins were mixed at 10 µM in 20 mM Na_2HPO_4 pH 7.4 at 25°C for 3 hr. Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining.

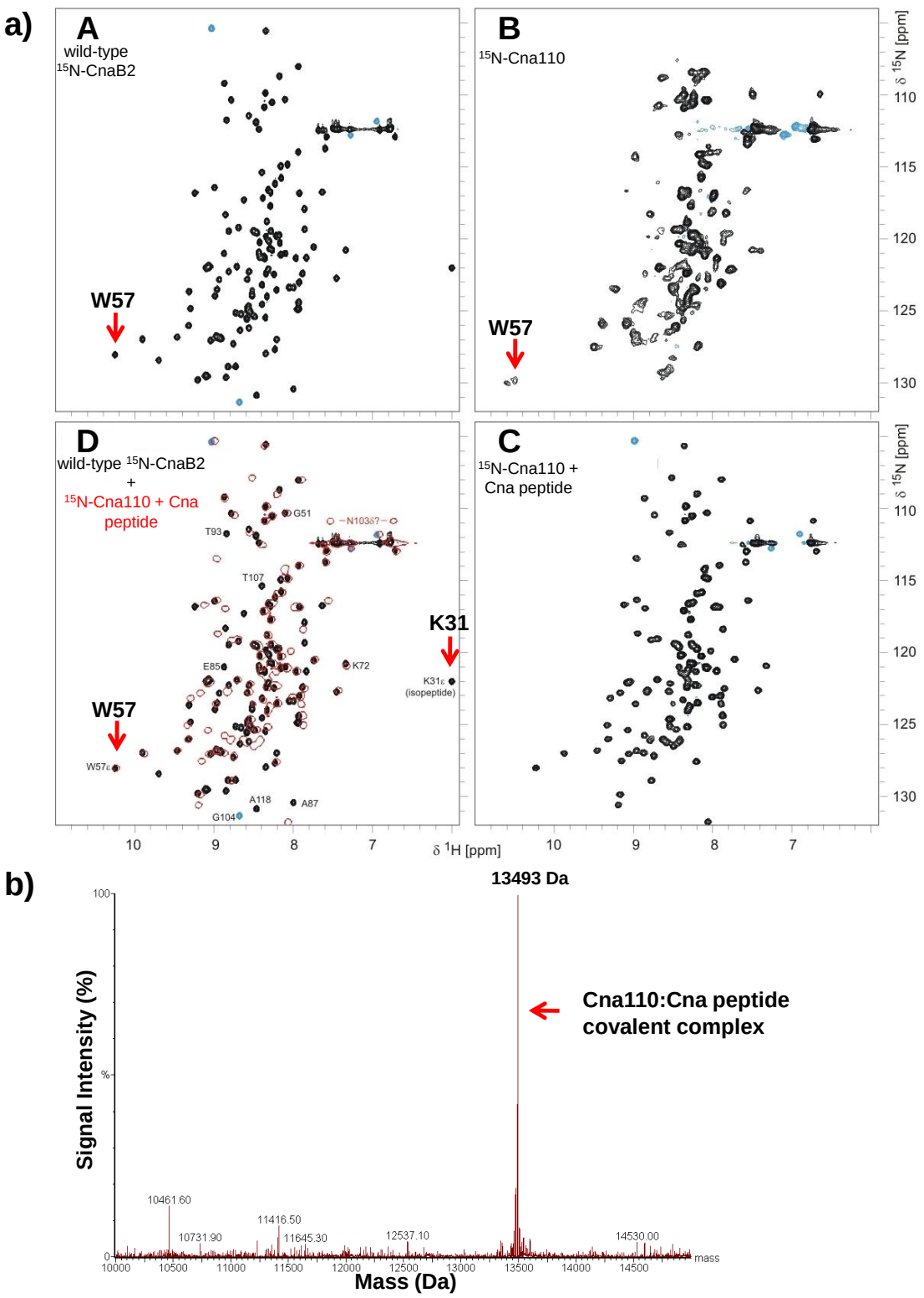


Figure 4.12. Analysis of ^{15}N-Cna110 and Cna peptide reconstitution with NMR spectroscopy and mass spectrometry. a) ^{1}H,^{15}N-HSQC spectrometric analysis of Cna110 reconstitution with Cna peptide. Box A: wild-type ^{15}N-CnaB2, Box B: ^{15}N-Cna110, Box C: ^{15}N-Cna110 + Cna peptide, Box D: overlay of wild-type ^{15}N-CnaB2 (black) and ^{15}N-Cna110 + Cna peptide (red) spectra. Red arrows point to key residues, where the W57 indole H-N signal and the K31 (isopeptide bond forming) signal are highlighted. b) Electrospray ionization mass spectrometry of the reaction between ^{15}N-Cna110 (12,054

Da) and Cna peptide (1,456 Da). Isopeptide bonded ^{15}N-Cna110 and Cna peptide has an expected mass of 13,492 Da due to the release of a water molecule (18 Da) (12,054 Da + 1,456 Da – 18Da). Note: ^{15}N-Cna110 and Cna peptide were prepared by Bijan Zakeri, however the NMR and mass spectrometry experiments in this figure were performed by Benjamin Schomburg in the Schwarz-Linek group.

4.3. Summary

In proof of principle studies to develop a covalently binding peptide tag, I split Spy0128 to develop a peptide-protein pair that reconstituted through spontaneous isopeptide bond formation. Two limitations of the isopeptag:pilin-C system were the large size of the binding partner and the inefficiency of the reconstitution reaction. I began to address these concerns in this chapter by exploring other protein scaffolds for the development of similar technologies. Using structural similarity searches of the Protein Data Bank with an isopeptide bond triad template I was able to find ACE19, a protein that contained a potential intramolecular isopeptide bond and was approximately half the size of Spy0128. Others have also identified ACE19 as containing a previously unrecognized intramolecular isopeptide bond, however the presence of the bond had not been experimentally confirmed (Kang et al., 2007; Kang and Baker, 2011; Zahner et al., 2011). After verifying the presence of the isopeptide bond with mass spectrometry, I split the protein to make a peptide and two protein binding partners. Although they were able to covalently reconstitute, the yield of the reactions were very low and therefore not an improvement on the isopeptag:pilin-C system. In my structural similarity searches, I came across a large number of proteins that had Lys, Asn/Asp, and Glu/Asp residues positioned in orientations that resembled those present in isopeptide bond containing proteins. As we gain a deeper understanding of the structural requirements of intramolecular isopeptide bond formation, it may be possible to use protein engineering techniques to modify these proteins to provide them with the ability to form spontaneous intramolecular isopeptide bonds. This approach could be used to provide enhanced

structural stability to various proteins and enzymes, which would have many practical applications.

I also investigated CnaB2 as a platform for developing covalent peptide-protein pairs. CnaB2 is a small protein domain that is almost one-third the size of Spy0128 and it contains one intramolecular isopeptide bond. I split CnaB2 to make the eight residue peptide PepA and two versions of a binding partner (Cna110 and Cna114). Due to the presence of 4 hydrophobic residues PepA was modified to make PepB, with the two Val residues mutated to Thr to make the peptide more hydrophilic. PepA was much more efficient at reconstitution than PepB, and reconstitution with Cna110 resulted in a slightly higher product yield than with Cna114. Therefore, the best reconstituting pair was PepA and Cna110. Preliminary characterization experiments showed that PepA and Cna110 reconstitution proceeded to a similar extent at a range of temperatures and was more efficient than the isopeptag:pilin-C pair, where almost one-quarter of Cna110 was able to reconstitute within the first few seconds of reaction. Another attractive feature of PepA and Cna110 reconstitution was that there did not appear to be any detectable side-reactions occurring unlike the isopeptag:pilin-C pair. Therefore, the PepA:Cna110 pair represented a small and efficiently reconstituting peptide-protein pair.

For practical applications of this technology it would be ideal to have a peptide-protein pair that could rapidly reconstitute with a high product yield. Although the PepA:Cna110 system was more efficient than the isopeptag:pilin-C system, it may have still been possible to further improve reconstitution efficiency. Therefore, I had to explore ways of improving the efficiency of reaction for the PepA:Cna110 pair. To do so, *in vitro* protein evolution techniques such as phage display or protein engineering could be used. Using phage display I could search a large peptide library to find an efficient covalently binding peptide for the split CnaB2 constructs. Alternatively, I could use protein engineering techniques to make a series of mutants to try to enhance the binding properties of the peptide and its protein partner.

Initially, I hypothesized that the split protein fragments would either be present in a partially unfolded state, or that they may adopt a new conformation that minimized the exposure of core hydrophobic regions that resulted from removal of a β-strand. To shed light on this matter we used NMR spectroscopy to explore the structural aspects of protein reconstitution. Our preliminary investigations revealed that the split Cna110 construct had a disordered structure compared to wild-type CnaB2. However, when the missing β-strand was added to the split protein, the peptide-protein pair was able to rapidly reconstitute in to a structure resembling the wild-type protein and form an intermolecular isopeptide bond as determined by mass spectrometry. However, we cannot draw definitive conclusions from these results until we perform more detailed structural investigations. Furthermore, based on these early findings it appeared that our split constructs could also be used to study the conformation of split proteins and the structural aspects of split protein reconstitution. Moreover, the proposed mechanism of isopeptide bond formation has only been studied in proteins where isopeptide bond formation has gone to completion. The split CnaB2 constructs would also allow for more detailed studies of the mechanism of spontaneous isopeptide bond formation.

Chapter 5:

Phage display and rational modifications to develop an efficiently reconstituting covalent peptide-protein pair

Chapter 5: Phage display and rational modifications to develop an efficiently reconstituting covalent peptide-protein pair

Exploration of three different intramolecular isopeptide bond containing proteins had led to the development of a small and efficient covalently reacting peptide-protein pair. These studies demonstrated the generality of the approach used for the development of covalent peptide fusion tags. Although the split CnaB2 system was smaller and more efficient than the split Spy0128 system, it may have still been possible to improve the split CnaB2 platform to develop a more efficiently reacting peptide-protein pair. This technology has the potential to have widespread relevance for a variety of biological applications. However, ideally the peptide-protein pair reaction needs to be more efficient. As such, I proceeded to try to optimize the interaction between the split CnaB2 constructs to achieve more efficient covalent reconstitution.

Certain outer-membrane proteins of Gram-positive bacteria have been optimized by nature to form intramolecular isopeptide bonds, and these bonds form rapidly as the proteins naturally fold (Hu et al., 2010). I have modified three of these proteins to develop peptide-protein pairs that interact through intermolecular isopeptide bond formation, and it is not surprising that the intermolecular bond formation is less efficient than its intramolecular counterpart. To improve intermolecular peptide-protein interactions, I attempted to optimize the reactants specifically for this function using phage display and rational modifications.

5.1. Phage display for isolating efficiently reacting covalent peptides

5.1.1. Introduction to phage display

5.1.1.1. Bacteriophage M13 phage display libraries

Phage display is an elegant technique for screening a large peptide or protein library against a target molecule of interest (Smith, 1985). This technique allows for the

display of a diverse polypeptide library on the surface of phage as fusion constructs of phage coat proteins, and often a library diversity of 10^9-10^{10} different variants can be achieved (Kehoe and Kay, 2005). The most common phage used for phage display is the filamentous bacteriophage M13 (Fig. 5.1a) (Kehoe and Kay, 2005). The circular single stranded DNA (ssDNA) of M13 is packaged by 2700 copies of the major coat protein pVIII to form a long cylindrical structure that is capped at both ends by several minor coat proteins, one of which is pIII and is present in 5 copies (Fig. 5.1a) (Kehoe and Kay, 2005). The proteins pIII and pVIII constitute the most commonly used coat proteins of M13 for the display of polypeptide libraries as fusion constructs (Kehoe and Kay, 2005). Since pVIII is present in such large quantities, peptide libraries fused to pVIII are more applicable to isolating low affinity binders due to the avidity effect (Kehoe and Kay, 2005; Kay et al., 2001). On the other hand, the limited number of pIII molecules reduces the avidity effect and allows for the identification of high affinity ligands (Kehoe and Kay, 2005; Kay et al., 2001). To generate a randomized peptide or protein fusion library, the corresponding coding sequence with randomized positions is cloned as a fusion construct of a phage coat protein (for example pIII or pVIII) in the phage genome (Kehoe and Kay, 2005). Then when the phage propagate, the virions each express a different version of the peptide or protein under investigation as surface-displayed fusion constructs of phage proteins (Kehoe and Kay, 2005). This serves as a way to link phenotype to genotype, and therefore when a ligand with the desired properties is isolated from phage screens, the sequence of the displayed peptide or protein can be readily determined by sequencing the corresponding region on the phage genome.

5.1.1.2. General phage display panning protocol

While the theory behind a phage library screen is straightforward, the difficulty lies in customizing the protocol to meet the requirements of the experiment, in order to isolate a peptide or protein with the desired binding properties. Fig. 5.1b outlines a typical panning experiment for the screening of a peptide library against a target protein (New

England Biolabs, 2009). The first step prior to library screening is to tether the target protein to a solid support such as magnetic beads. Subsequently, the phage library is incubated with the target molecule under conditions and for the amount of time as dictated by the binding properties that are being sought (Fig. 5.1b). This is critical as one must be very selective of the experimental conditions used for the screen in order to isolate a peptide that has been optimized to bind under the desired conditions. Unbound phage are then washed away using buffer stringencies as dictated by the desired binding affinity (Fig. 5.1b). Finally, bound phage are eluted and amplified for further rounds of panning (Fig. 5.1b) (New England Biolabs, 2009). In subsequent rounds of panning, washing stringencies can be increased in order to enrich the phage pool with peptides that possess higher binding affinities for the target protein. After the desired number of rounds of panning has been performed, isolated phage are then individually screened to verify whether they possess the desired binding properties, and subsequently sequenced. Typically three to four rounds of panning is required to obtain a consensus sequence for the optimized peptide based on the manufacturer's recommendations (New England Biolabs, 2009).

Herein, my objectives were to use phage display to isolate a peptide that could covalently bind the split CnaB2 binding partner more efficiently than PepA through intermolecular isopeptide bond formation. Since I was only interested in covalently binding peptides, I used highly stringent conditions in my panning experiments in order to reject high affinity non-covalently binding peptides. To conduct the phage display experiments, I used the commercially available dodecapeptide phage library (Ph.D.-12^{TM}) from New England Biolabs, where a linear randomized 12 residue peptide library with a diversity of 10^9 clones was fused to the pIII minor coat protein of M13 phage (New England Biolabs, 2009).

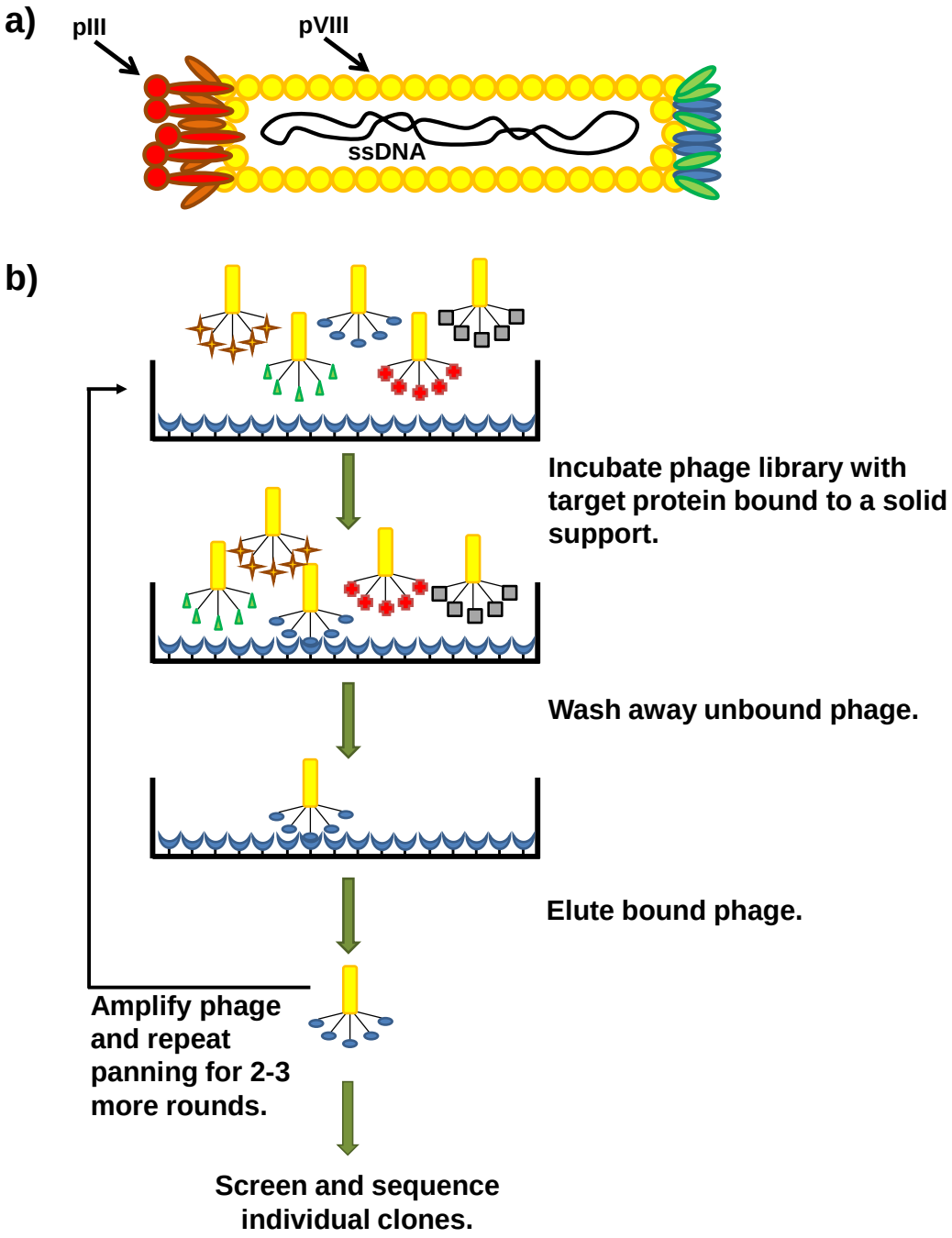


Figure 5.1. Phage display. a) A cartoon of M13 bacteriophage. The ssDNA genome of M13 is packaged inside a cylindrical coat composed of 2700 copies of the major coat protein pVIII, and capped at each end by minor coat proteins. The minor coat protein pIII is present in 5 copies at one end of the phage and is commonly used as the fusion partner to display peptide or protein libraries. This figure was adapted from (Kehoe and Kay, 2005). b) A typical phage display panning experiment of a peptide library. The phage library is incubated with a target protein (bound to a solid support) to allow for binding. Unbound phage are then washed and the bound phage are eluted for further rounds of panning. After 2-3 more rounds of panning, the enriched pool of phage are individually screened and sequenced to obtain a consensus sequence for an optimized binding peptide. This figure was adapted from (New England Biolabs, 2009).

5.1.2. Phage display screens for efficient and covalently binding peptides: Experimental setup

Prior to starting the phage display library screens, I had to tether Cna110/Cna114 (both protein binding partners were tested) to a solid support. The objectives of these screens were to identify covalently binding peptides to Cna110/Cna114, and thus stringent wash steps were required in each round of panning to disrupt all non-covalently binding peptides. Consequently, Cna110/Cna114 had to be secured to a solid surface via a stable covalent linkage to ensure the protein would not dissociate during panning experiments. Furthermore, Cna110/Cna114 was a fragmented protein in a partially disordered state and it would be advantageous for the protein to have maximum flexibility in solution to mediate efficient reconstitution. Accordingly, it was decided that the best strategy was to covalently link Cna110/Cna114 to magnetic beads. This allowed Cna110/Cna114 to have more flexibility in solution as opposed to adhering the protein to the bottom of a 48-well plate via non-covalent interactions for example, which may have caused Cna110/Cna114 to adopt conformations not amenable for reconstitution.

To link Cna110/Cna114 to magnetic beads, a Cys residue was introduced near the N-terminal 6X His tag of the recombinant protein to make Cys-Cna110/Cys-Cna114 (Fig. 5.2a). Cys-Cna110/Cys-Cna114 was then conjugated to magnetic beads via the cleavable cross-linker sulfo-LC-SPDP (Thermo Scientific). To conjugate Cys-Cna110/Cys-Cna114 to magnetic beads, amine coated magnetic beads were activated with sulfo-LC-SPDP via the reactive *N*-hydroxysuccinimide (NHS) ester moiety on the cross-linker (Fig. 5.2b) (Thermo Scientific, 2008). Subsequently, the activated magnetic beads were coated with Cys-Cna110/Cys-Cna114 via reaction of the sulfhydral group on the introduced N-terminal Cys residue with the cross linker resulting in the release of pyridine 2-thione (Fig. 5.2b) (Thermo Scientific, 2008). To elute Cys-Cna110/Cys-Cna114 from the magnetic beads, DTT could be added to reduce the disulfide bond between the Cys residue and the cross-linker (Fig. 5.2b) (Thermo Scientific, 2008). Based on manufacturer recommendations, I initially used DTT to elute Cys-Cna110/Cys-Cna114

from magnetic beads. However, a concern was that Cys containing phage in the library may form disulfide linkages with the magnetic beads, whereby DTT elution would also result in the release of phage that did not interact with Cys-Cna110/Cys-Cna114. Therefore, as an alternative elution method Cys-Cna110/Cys-Cna114 could also be eluted from magnetic beads by protease cleavage using the Tobacco Etch Virus (TEV) protease cleavage site, between the introduced Cys residue and Cna110/Cna114 in the recombinant construct (Fig. 5.2a). Using a similar approach Cys-CnaB2 and Cys-Cna110 E77Q, which had the catalytic Glu^{77} mutated to prevent isopeptide bond formation, were made as negative controls for the screens to assess non-covalent phage binding.

For panning experiments, I tested many different conditions to isolate covalently binding peptides. My general procedure for a typical panning experiment was to mix the magnetic beads conjugated to CnaB2 based constructs with the Ph.D.-12^{TM} phage library, and incubate the mixture at 25°C for 5-16 hr in the first round of panning. These conditions were chosen in order to select for peptides that had optimal binding at 25°C, and to allow sufficient time for reaction in order to isolate the maximum amount of covalently binding peptides. In subsequent panning cycles, the objectives were to reduce the time of incubation of the library with Cys-Cna110/Cys-Cna114 in order to separate the more efficiently binding peptide. Next, several harsh washes were performed to try to remove all non-covalently reconstituting peptides from the Cys-Cna110/Cys-Cna114-conjugated beads. In doing so, I found that phage were generally very resilient and remained viable under most of the harsh conditions to which they were exposed. Subsequently, DTT or TEV protease cleavage was used to elute Cys-Cna110/Cys-Cna114 and any phage that were bound to the binding partners, and then the eluate was titred.

Each round of panning was validated to assess the efficiency of the panning protocol for isolating covalently binding phage, while eliminating non-covalently binding phage. Also, each round of panning took approximately one week to perform and in the interests of time it was important to validate the protocol at an early stage. To screen the

eluate for covalently binding phage, I would randomly select and amplify individual phage using plaques picked from the titration plates. Then the amplified phage were incubated with Cna110 and the reactions were analyzed using Western blot analysis with an antibody against pIII (the peptide library is fused to pIII). If Cna110 reacted covalently with any of the phage, then a noticeable mobility shift of the pIII band on the Western blot films was expected. If a mobility shift of the bands were not observed, then the isolated phage were either high affinity non-covalent binders or non-specific phage, and the first round of panning was repeated under more stringent conditions to eliminate non-covalently binding phage. The negative controls (Cys-CnaB2/Cys-Cna110 E77Q) were tested alongside the binding partners in the panning experiments to allow for the comparison of the amount of phage eluted from the screen, compared to the amount eluted as a result of non-covalent binding to the negative controls.

5.1.3. Phage display screens for efficient and covalently binding peptides: Results

A summary of the phage display experiments that I performed is provided in Table 5.1. Each round of panning (100 µL reactions) started with incubation of the binding partner with the phage library at a concentration of 1×10^{11} plaque forming units (pfu) per mL, and hence a total of 1×10^{10} pfu of phage were used for each screen. The initial round of screening was performed using manufacturer recommended washing conditions. Eluted phage were then tested using Western blot analysis. However, these eluted phage failed to yield any covalently binding peptides (Table 5.1). A representative Western blot film of a typical screen is shown in Fig. 5.2c where there was no observable shift in the mobility of pIII, and other Western blots performed gave similar results. Consequently, in subsequent panning experiments I began to gradually increase the stringency of the wash steps to eliminate non-covalently binding phage (Table 5.1). The phage were very resilient and able to tolerate all of the harsh washing conditions tested except for a 10 min wash in PBS containing 3 M Guanidinium hydrochloride (GdmCl). Regardless of the stringency of the washing conditions tested, all of the phage isolated and screened by

Western blot analysis were not able to covalently bind Cna110 (Table 5.1). Overall, I tested the phage library with the Cys-Cna110 and Cys-Cna114 platforms bound to magnetic beads and even an ELISA plate in one experiments, and explored many harsh wash conditions including high detergent, high salt, low pH, and high temperature (Table 5.1). Yet, I was unable to find any covalently binding peptides (Table 5.1).

Since I only screened a small number of phage from each panning experiment, there is the possibility that I simply missed the covalently binding phage that may have been isolated in the screens. Individual screening of phage was a low-throughput and labour intensive process and thus for each experiment it was only feasible to screen a small subset of phage. However, in all of the experiments performed I was unable to find a single covalently binding peptide regardless of whether it was more or less efficient at reconstitution than PepA. A possibility for this may have been that the split CnaB2 proteins were not tolerant to mutations in the sequence of the missing peptide (investigated in section 5.2.5.), thus reducing the likelihood of finding a covalently binding peptide in a random library. Alternatively, a covalently binding peptide may not have been present in the library as it had a diversity in the order of 10^9 clones, which does not cover all of the sequence complexity possible in a 12 residue peptide (New England Biolabs, 2009). Furthermore, an inherent limitation of these experiments was that the M13 phage is an assembly of proteins held together through non-covalent interactions, and if I were to test conditions that eliminated all non-covalent interactions it would likely kill all of the phage. This was observed in an experiment where I used 3 M GdmCl in one of the wash steps (Table 5.1).

In the phage display experiments performed, I did not have an appropriate control for a peptide-displaying phage that could bind Cys-Cna110/Cys-Cna114 covalently, and thus I could not validate the reliability of my panning and screening protocols for isolating a covalently binding phage. In future experiments, a phage expressing a pIII-PepA fusion construct must be developed. The pIII-PepA expressing phage could then be used to optimize the phage display protocol and demonstrate that covalently binding phage can

be isolated from a pool of random phage clones and detected using Western blot analysis. Subsequently, a focused phage library could be produced where the residues necessary for isopeptide bond formation are preserved but surrounding residues are randomized. This would allow for a more direct approach for isolating an efficiently binding covalent peptide.

Considering the complications of adapting phage display for my unique requirements and the lack of isolating any covalently binding phage, I decided to use a more direct approach to find more efficiently reacting peptides for the split CnaB2 constructs. By exploring the chemical diversity in the binding interface between the peptide and protein, I used rational modifications to try to introduce new interactions that may enhance binding.

Experiment	Conjugated Binding Partner	Binding Conditions	Wash Steps	Eluting Technique	Eluate Titre	Western Blot Results	Comments
1	- Cys-Cna110	25°C for 6 hr at 1400 rpm	i) 3 x Tris-Buffered Saline (TBS: 50 mM Tris-HCl + 150 mM NaCl) + 0.1% Tween20 pH 7.5 ii) 1 x 0.2 M Glycine-HCl pH 2.2 for 20 min at 25°C and 1400 rpm iii) 3 x TBS + 0.5% Tween20 pH 7.5	DTT	- Cys-Cna110 = $4.8x10^5$ pfu/mL	Negative	Six screens were also sequenced and peptide sequences were random.
2 (Panning Round 1)	- Cys-Cna110	25°C for 6 hr at 1400 rpm	i) 3 x TBS + 0.1% Tween20 pH 7.5 ii) 2 x 0.2 M Glycine-HCl pH 2.2 for 10 min at 25°C and 1400 rpm iii) 3 x TBS + 0.5% Tween20 pH 7.5	DTT	- Cys-Cna110 = $8.3x10^7$ pfu/mL	Negative	Use eluate to perform a second round of panning.
2 (Panning Round 2)	- Cys-CnaB2 - Cys-Cna110	25°C for 30 min at 1400 rpm	i) 3 x TBS + 0.1% Tween20 pH 7.5 at 65°C and 1400 rpm ii) 2 x 0.2 M Glycine-HCl pH 2.2 for 10 min at 25°C and 1400 rpm iii) 3 x TBS + 0.5% Tween20 pH 7.5	DTT	- Cys-CnaB2 = $3.9x10^6$ pfu/mL - Cys-Cna110 = $2.7x10^6$ pfu/mL	Not performed	The negative control had a higher titre than Cna110. Therefore, try binding proteins to an ELISA plate.
3	- Cys-CnaB2 - Cys-Cna110 *Both coated on an ELISA plate.	25°C for 5 hr on rocker	i) 3 x TBS + 0.1% Tween20 pH 7.5 ii) 3 x TBS + 0.1% Tween20 pH 7.5 for 5 min at 60°C iii) 2 x PBS pH 7.5 + 0.9 M NaCl for 10 min at 25°C iv) 3 x TBS + 1% Tween20 pH 7.5 for 2 min	N/A (phage were amplified directly on plate)	- Cys-CnaB2 = $1.4x10^4$ pfu/mL - Cys-Cna110 = $3.3x10^4$ pfu/mL	Not performed	There was still too many phage from the negative control. ELISA plate did not provide an advantage, use beads but elute with TEV protease and use stringent washes.
4	- Cys-CnaB2 - Cys-Cna110	25°C for 6 hr at 1400 rpm	i) 4 x TBS + 0.5% Tween20 pH 7.5 ii) 1 x TBS + 0.5% Tween20 pH 7.5 for 10 min at 65°C and 1400 rpm iii) 1 x PBS pH 7.4 + 0.9 M NaCl for 10 min at 25°C and 1400 rpm iv) 1 x 0.2 M Glycine-HCl pH 2.2 for 10 min at 25°C and 1400 rpm v)1 x PBS pH 7.5 + 3 M GdmCl for 10 min at 25°C and 1400 rpm vi) 3 x TBS + 0.5% Tween20 pH 7.5	TEV protease	- Cys-CnaB2 = 0 plaque - Cys-Cna110 = 0 plaque	N/A	Phage did not remain viable likely due to a wash containing 3 M GdmCl. Therefore, repeat without washing with 3 M GdmCl.

5	- Cys-CnaB2 - Cys-Cna110	25°C for 6 hr at 1400 rpm	i) 4 x TBS + 0.5% Tween20 pH 7.5 ii) 1 x TBS + 0.5% Tween20 pH 7.5 for 10 min at 65°C and 1400 rpm iii) 1 x PBS pH 7.5 + 0.9 M NaCl for 10 min at 25°C and 1400 rpm iv) 1 x 0.2 M Glycine-HCl pH 2.2 for 10 min at 25°C and 1400 rpm v) 3 x TBS + 0.5% Tween20 pH 7.5	TEV protease	- Cys-CnaB2 = $4.0x10^3$ pfu/mL - Cys-Cna110 = $8.0x10^3$ pfu/mL	Negative	Phage did not amplify well. Repeat experiment.
6	- Cys-CnaB2 - Cys-Cna110	25°C for 6 hr at 1400 rpm	i) 4 x TBS + 0.5% Tween20 pH 7.5 ii) 1 x TBS + 0.5% Tween20 pH 7.5 for 10 min at 65°C and 1400 rpm iii) 1 x PBS pH 7.5 + 0.9 M NaCl for 10 min at 25°C and 1400 rpm iv) 1 x 0.2 M Glycine-HCl pH 2.2 for 10 min at 25°C and 1400 rpm v) 3 x TBS + 0.5% Tween20 pH 7.5 vi) 1 x PBS pH 7.5	TEV protease	- Cys-CnaB2 = $1.8x10^6$ pfu/mL - Cys-Cna110 = $5.7x10^6$ pfu/mL	Negative	Cna110 as the binding partner has not produced any results. Therefore, try using Cna114 as the binding partner and Cna110 E77Q as a negative control.
7	- Cys-Cna110 E77Q - Cys-Cna114	25°C for 16 hr at 1200 rpm	i) 3 x TBS + 0.5% Tween20 pH 7.5 ii) 1 x TBS + 1% Tween20 pH 7.5 for 10 min at 65°C and 1400 rpm iii) 1 x PBS pH 7.5 + 0.9 M NaCl for 10 min at 65°C and 1400 rpm iv) 1 x 0.2 M Glycine-HCl pH 2.2 for 10 min at 65°C and 1400 rpm v) 3 x TBS + 1% Tween20 pH 7.5 vi) 1 x PBS pH 7.5	TEV protease	- Cys-Cna110 E77Q = 0 plaques - Cys-Cna114 = $9.0x10^3$ pfu/mL	Negative	This experiment gave the best results with 0 phage isolated from the negative control, but screening of the isolated phage by Western blot analysis did not demonstrate any covalent binding.

Table 5.1. A summary of phage display experiments. This table provides a summary of the various binding proteins, conditions, and approaches that were explored in the phage display panning experiments and their corresponding results. Cys-Cna110 and Cys-Cna114 were used as platforms to screen the phage peptide library against and Cys-CnaB2 and Cys-Cna110 E77Q were used as negative controls for the screens to assess non-covalent binding.

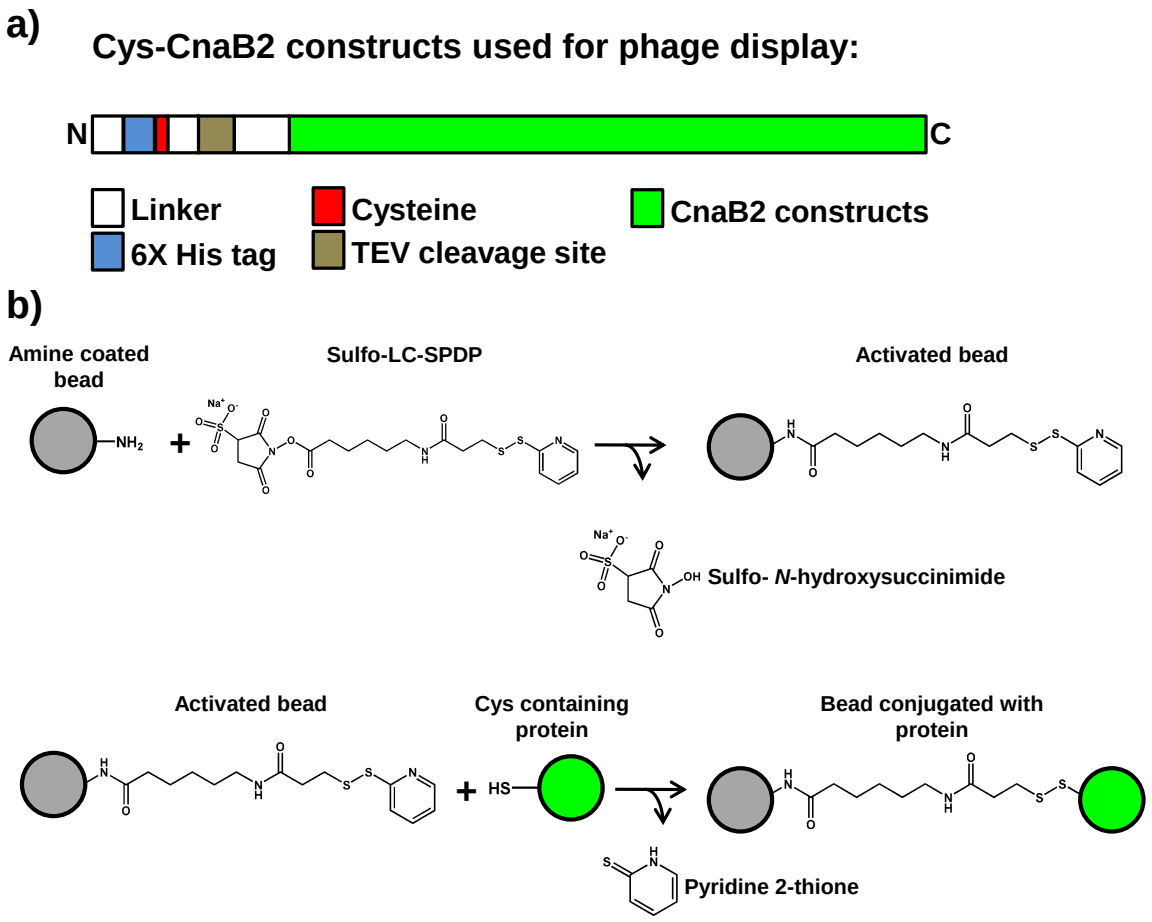


Figure 5.2. Phage display approaches. a) A map of Cys containing CnaB2 based constructs (Cys-CnaB2, Cys-Cna110, Cys-Cna114, Cys-Cna110 E77Q) used for phage display panning experiments, with a legend of relevant sites in the recombinant constructs. A Cys residue was introduced near the N-terminal 6X His tag, which was used to tether various CnaB2 constructs to magnetic beads via a cross-linker (that contained a disulfide bond which could be cleaved using a reducing agent). Phage bound to Cna110/Cna114 were eluted from

the magnetic beads using either DTT reduction or TEV protease cleavage. b) Conjugation of amine coated magnetic beads with Cys containing CnaB2 constructs. Beads were first activated with the sulfo-LC-SPDP cross-linker which released sulfo-*N*-hydroxysuccinimide. Activated beads were then conjugated to Cys containing proteins via reaction of the Cys thiol group with the cross-linker, which released pyridine 2-thione. c) A typical Western blot result from a screen for phage display panning experiments using an anti-pIII antibody. This was a Western blot film from experiment 2 (panning round 1) shown in Table 5.1. Reactions were prepared by mixing 5 µL of Cna110 (151 µM) with 10 µL of amplified phage (typically at 10^{13} pfu/mL) in 5 µL 20 mM Na_2HPO_4 pH 7.4. The reactions were incubated for 6 hr at 25°C, and then heated to 95°C for 7 min in SDS-loading buffer containing 50 mM DTT and 12 µL of each reaction was run on 10% SDS-PAGE followed by Western blot analysis. Amplified M13 represents a stock solution of M13 (typically at 10^{13} pfu/mL) used as a positive control for a pIII signal (prepared as above except with 5 µL of amplified phage in 15 µL 20 mM Na_2HPO_4 pH 7.4). Cys-Cna110 is the purified protein used as a negative control. pIII-peptide fusion has an expected molecular mass of ~42 kDa but on a gel it gives a band at ~60 kDa, which is consistent with past observations (Tesar et al., 1995). If the pIII-peptide fusion formed a covalent bond to Cys-Cna110, then a band at ~75 kDa would be expected. *Unknown bands that may have been a result of pIII-peptide fusion degradation.
Figure 5.2b was adapted from the sulfo-LC-SPDP manual (Thermo Scientific) (Thermo Scientific, 2008).

5.2. Rational modification experiments for isolating efficiently reacting covalent peptide-protein pairs

5.2.1. Rational modifications of Cna114 and PepA

The phage display experiments that I performed failed to yield any leads for a more efficiently reacting covalent peptide-protein pair. Therefore, rather than searching a random peptide library it was decided to rationally modify the covalent peptide-protein pair to enhance reconstitution efficiency. Since I was working towards engineering a more efficient interaction between the split CnaB2 constructs, I decided to conduct all mutagenesis experiments on Cna114 and the most efficient binding peptide PepA, which had the wild-type CnaB2 sequence. While Cna110 was slightly more efficient at reconstitution than Cna114 (Fig. 4.8b), our NMR experiments demonstrated that Cna110 was in a partially disordered state (Fig. 4.12a). While similar studies have not been performed on Cna114, it was hypothesized that the extra four residues in Cna114 may help to provide a more stable structure than Cna110. Therefore, it was concluded that Cna114 would be the better platform to attempt to improve through mutagenesis.

5.2.2. Approaches taken to optimize Cna114 reconstitution efficiency

To optimize the efficiency of reconstitution of Cna114 I explored three different approaches (Fig. 5.3). One approach was to stabilize the fragmented Cna114 protein by increasing the C-terminal end of the split protein by 1 and 2 residues to create Cna115 and Cna116 (Fig. 5.3a). The rationale for this was that as the fragmented protein is extended in to the C-terminal β-strand, the split protein may be better able to form a structure resembling wild-type CnaB2 with an 'open' binding pocket for peptide binding. Alternatively, I sought to enhance the peptide-protein binding efficiency by exploring the chemical space around the C-terminal β-strand. In CnaB2, there is a flexible loop that is close to the C-terminal β-strand and the residue Ile^{34} in the loop may be interacting with Val^{116} in the β-strand (Fig. 5.3b). Although this interaction was not clear in the crystal structure, it was hypothesized that the loop may adopt alternate conformations in solution, and in the fragmented Cna114 protein the loop may interact with core hydrophobic residues and reduce reconstitution efficiency of the peptide. Therefore, to reduce potential hydrophobic interactions between Ile^{34} and the hydrophobic core of the split protein, I investigated mutating Ile^{34} to the more polar residues Thr, Arg, Asp, Asn, Glu, and Ser (Fig. 5.3b). These residues allowed analysis of polar, negatively charged, and positively charged residues at position 34. The third approach that I undertook for optimizing Cna114 as a binding partner was to try to increase the solubility of the protein by mutating surface-exposed hydrophobic residues to hydrophilic ones. CnaB2 is a domain of the much larger protein FbaB, which contains another CnaB domain and several disordered sequence repeats (Hagan et al., 2010). I hypothesized that within the complete FbaB protein, the CnaB2 domain is likely to interact with its surrounding domains and it is possible that when CnaB2 is in solution it may make similar transient interactions with other CnaB2 molecules. These transient interactions may reduce the reconstitution efficiency of the split CnaB2 constructs. Therefore, to increase the hydrophilic nature of Cna114 I mutated two surface-exposed hydrophobic residues (Fig. 5.3c). Met^{69} and Ile^{58} were solvent-exposed

hydrophobic residues that were on the same face of CnaB2 and I mutated these residues to Tyr, a more hydrophilic residue that was also favoured in β-sheets (Otaki et al., 2010).

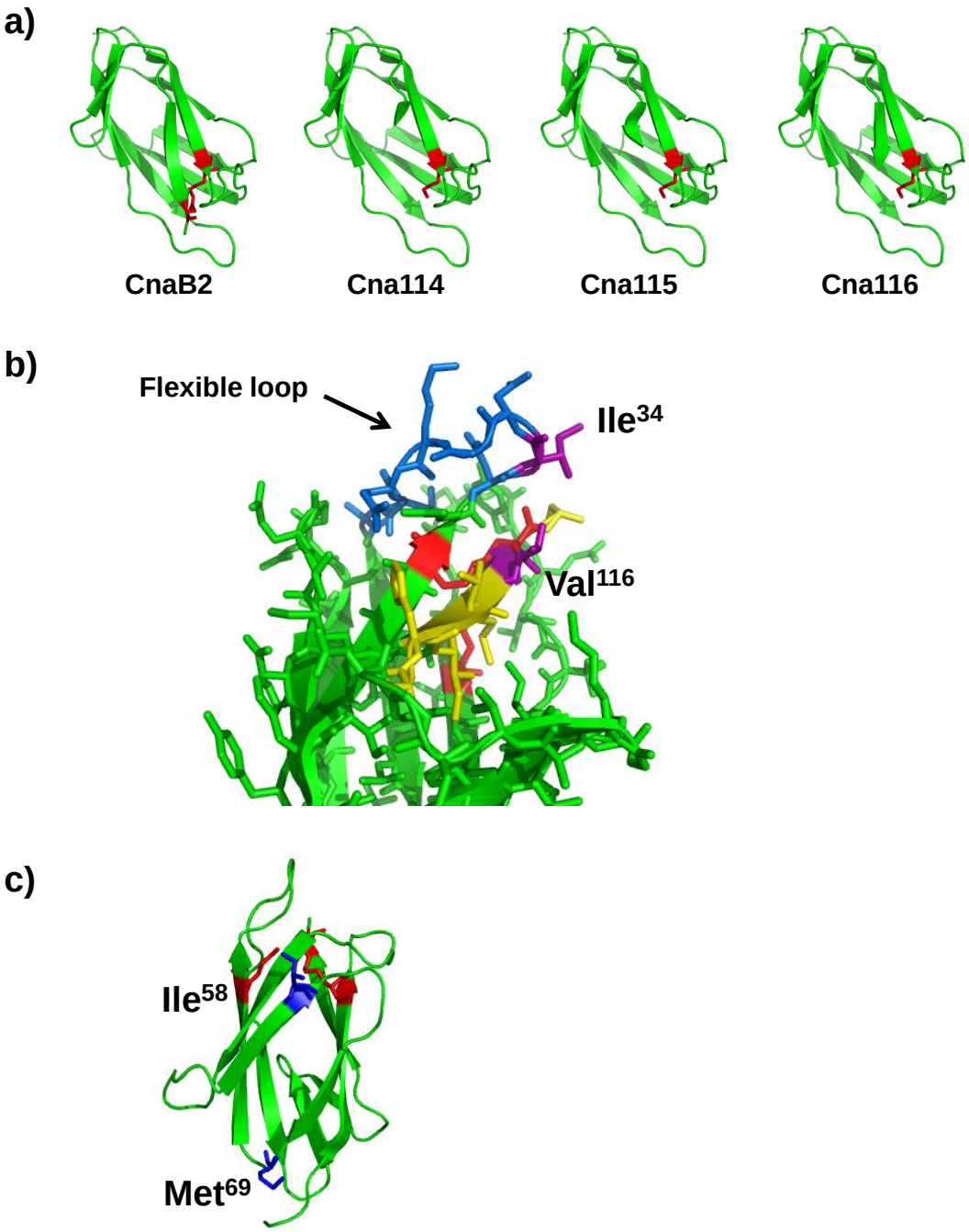


Figure 5.3. Optimization of the Cna114 platform. a) To attempt to stabilize the fragmented Cna114 protein, Cna114 was extended by 1 and 2 residues at its C-terminal end to make Cna115 and Cna116. Isopeptide bond forming residues are shown in red. b) The flexible loop

(blue) in CnaB2 may interact with the C-terminal β-strand (yellow). Particularly, the hydrophobic Ile^{34} and Val^{116} residues (purple) may be interacting in solution. Therefore, to explore the nature of this interaction Ile^{34} was mutated to the more polar residues Thr, Arg, Asp, Asn, Glu, and Ser. Isopeptide bond forming residues are shown in red. c) CnaB2 may be forming transient hydrophobic interactions with other CnaB2 molecules in solution. Ile^{58} and Met^{69} (blue) which are hydrophobic residues present on the same face of the protein were mutated to the polar residue Tyr in an attempt to increase the hydrophilic nature of Cna114. Isopeptide bonding residues are shown in red. (PDB 2X5P)

5.2.3. Effects of Cna114 modifications on reconstitution efficiency

To attempt to enhance the reconstitution efficiency of Cna114, I cloned and expressed all of the mutants outlined in Fig. 5.3. The constructs all expressed well and purified as soluble proteins similar to the parental Cna114 protein. The reconstitution efficiency of the modified Cna114 binding partners were tested with CnaB2 peptides and the results are summarized in Table 5.2. Reconstitution efficiency was evaluated qualitatively using SDS-PAGE analysis to test if the modifications made to Cna114 were beneficial and representative examples of these gels are provided in the appendix in Fig. A1 and Fig. A2. Increasing the size of Cna114 by 1 residue to make Cna115 did not have a noticeable effect on reconstitution efficiency (Table 5.2). However, when two residues were added to make Cna116 the binding partner lost all activity (Table 5.2). Cna116 stopped two residues prior to the reactive Asp in the C-terminal β-strand of CnaB2 (Fig. 5.3a). It was likely that Cna116 was able to fold in to a structure that resembled CnaB2 and lacked the disordered structure observed for Cna110 using NMR spectroscopy. Thus, the binding site for the peptide was likely sterically hindered and the eight residue peptide could not bind to and covalently react with the binding partner.

To enhance the hydrophilic character of a flexible loop near the C-terminal β-strand (Fig. 5.3b), I investigated a series of mutations to Ile^{34}. The mutations made were intended to explore the chemical space around position 34 by introducing amino acids with polar, positively charged, and negatively charged side chains. When Arg was introduced at position

34 there was no significant change in reconstitution efficiency (Table 5.2). Interestingly, while introduction of the polar residues Thr and Asn did not have an effect on Cna114 reconstitution efficiency, the polar residue Ser did slightly improve the activity of Cna114 (Table 5.2). Furthermore, when the negatively charged residue Asp was introduced at position 34 it did not alter Cna114 activity; however the introduction of Glu at that position significantly improved reconstitution efficiency (Table 5.2). One possible reason for this effect may be that the surface-exposed hydrophobic residue Ile^{34} may have folded back on to the exposed groove of where the C-terminal β-strand was, and interacted with the hydrophobic core of the protein. A Glu at position 34 in the loop may have increased the hydrophilic character of the loop and led to a more accessible binding pocket for the peptide. Therefore, an I34E mutation in Cna114 was the best mutation investigated at position 34 and it led to improved reconstitution efficiency.

To increase the hydrophilic character of Cna114, I mutated two surface exposed hydrophobic residues (Ile^{58} and Met^{69}) to Tyr (Fig. 5.3c). When compared to Cna114, the I58Y mutant had a similar level of activity and did not show a noticeable improvement on reconstitution (Table 5.2). However, the M69Y mutation made significant improvements to the reactivity of Cna114 (Table 5.2). This was very interesting as Met^{69} was at the opposite end of CnaB2 relative to the intramolecular isopeptide bond (Fig. 5.3c), yet it had a dramatic effect on intermolecular reconstitution. Therefore, Met^{69} may have contributed to transient hydrophobic interactions between Cna114 molecules, whereas Ile^{58} may not have been involved in such interactions.

Construct	Reason for alteration to Cna114 platform	Activity
Cna114	N/A.	+++
Cna115	To increase protein stability.	+++
Cna116	To increase protein stability.	0
Cna114 I34T	Explore the effect of a polar side chain at residue position 34.	+++
Cna114 I34R	Explore the effect of a positively charged side chain at residue position 34.	+++
Cna114 I34D	Explore the effect of a negatively charged side chain at residue position 34.	+++
Cna114 I34N	Explore the effect of a polar side chain at residue position 34.	+++
Cna114 I34E	Explore the effect of a negatively charged side chain at residue position 34.	+++++
Cna114 I34S	Explore the effect of a polar side chain at residue position 34.	++++
Cna114 I58Y	Increase hydrophilic character of Cna114.	+++
Cna114 M69Y	Increase hydrophilic character of Cna114.	+++++

Table 5.2. Summary of the results obtained from a series of mutations made to Cna114. The split CnaB2 binding partner Cna114 was used as a platform for mutagenesis to improve its reactivity. This table is a summary of extensive mutagenesis experiments that were evaluated using SDS-PAGE. Representative gels of experiments that tested the activity of Cna114 versus its mutants when reacted against a peptide tag are available in the appendix Fig. A.1 and Fig. A.2. Mutants were tested against either PepA-MBP or PepL-MBP (explained in detail in section 5.2.5) because both Cna114 and PepA scaffolds were modified at the same time to find more efficiently reacting peptide-protein pairs. When a mutant with higher activity was found, it was used as the new scaffold to conduct further improvements.

5.2.4. Optimized Cna114 binding partner

In the series of mutations made to improve the reconstitution efficiency of Cna114, I found that the two mutations of I34E and M69Y made significant improvements to the activity of Cna114 (Table 5.2). These mutations were made to address separate concerns about the Cna114 scaffold: the I34E mutation was made to increase the hydrophilic character of a flexible loop, and the M69Y mutation was made to increase the hydrophilic character of Cna114 (Fig. 5.3b and c). Therefore, it was decided that the best Cna114 scaffold would incorporate both of these mutations and I consequently made Cna114 I34E M69Y, which was named SpyCatcher. Although the effects of the I34E and M69Y mutations were not additive

(explained in section 5.2.6.5 and Fig 5.7), each was made to address a different concern regarding the activity of Cna114. Therefore, it was decided that it would be advantageous to have both mutations in the optimized binding partner SpyCatcher.

5.2.5. Approaches taken to optimize PepA reconstitution efficiency

5.2.5.1. Rationale for PepD/E/F/G/H/I/O

In order to increase the reconstitution efficiency of PepA, I performed a series of mutations on the peptide to address various concerns. PepA (sequence: AH**IVMV**DA) had four hydrophobic residues at its centre that are shown in bold. These hydrophobic residues may have caused the peptide to interact with nearby hydrophobic surfaces thereby reducing display of the peptide in solution, or promoted the misfolding of proteins to which the peptide was fused. Therefore, I sought to increase the hydrophilic nature of the peptide and investigated the mutation of some of the hydrophobic residues to more hydrophilic ones. The residues Ile^{113} and Met^{115} have their side chain embedded in the protein core and therefore they were not ideal for mutagenesis as this would disrupt the hydrophobic core of the protein (Fig. 5.4). Alternatively, Val^{114} and Val^{116} have surface-exposed side chains and were more amenable for modification (Fig. 5.4). Earlier, I had investigated mutating both Val^{114} and Val^{116} residues to Thr (PepB) and this led to reduced reconstitution efficiency (Fig. 4.8b). Therefore, I investigated mutating each Val residue individually to polar or charged (positively and negatively) residues to gain a deeper understanding of the binding interactions. Accordingly, I made a series of mutants (PepD/E/F/G/H/I/O) that evaluated the introduction of Thr, Lys, Glu, or Tyr at individual Val positions (Table 5.3).

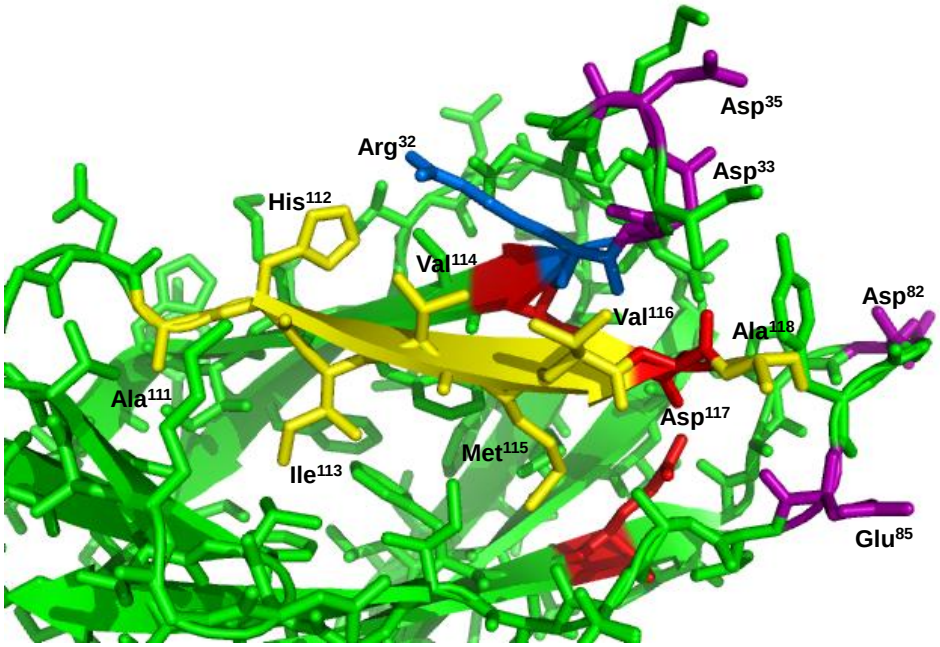


Figure 5.4. A close view of the C-terminal β-strand in CnaB2. An examination of the environment of PepA in CnaB2 that was investigated in the mutagenesis experiments to improve the binding efficiency of PepA to its split CnaB2 partner. CnaB2 is shown in green with the residues corresponding to PepA shown in yellow, isopeptide bond forming residues are shown in red, Arg^{32} is shown in blue, the negatively charged residues Asp^{33}, Asp^{35}, Asp^{82}, and Glu^{85} are shown in purple, and all of the relevant residues for mutagenesis investigations have been labelled. (PDB 2X5P)

Peptide (Pep-)	Sequence	Activity
A	AHIVMVDA	+++
B	AHITMTDA	+
C	RSGAHIVMVDAGSR	++++
D	AHITMVDA	++
E	AHIVMTDA	++
F	AHIKMVDA	+
G	AHIVMKDA	+
H	AHIEMVDA	+
I	AHIVMEDA	+
J	RSGAHIVMVDA	+++
K	AHIVMVDAGSR	++++
L	AHIVMVDAYKP	+++++
M	GDAHIVMVDA	++
N	AEIVMVDA	+
O	AHIVMYDA	++
P	AHIVMVDAYKPTK	+++++
P D117A	AHIVMVAAYKPTK	0

Table 5.3. A summary of the PepA variants and their activities. A series of peptide mutants based on the PepA scaffold were made and their activities were tested against the

optimized binding partner SpyCatcher. Residues altered in each mutant relative to PepA are shown in red. The activities are qualitatively shown relative to PepA and the quantified data is shown in Fig. 5.6b.

5.2.5.2. Rationale for PepC/J/K

In the earlier investigations of Cna110 and Cna peptide reconstitution, I observed rapid reconstitution of the free peptide with Cna110 (Fig. 4.11b). This may have either been due to an excess amount of Cna peptide present in the reaction mixture or that Cna peptide was more reactive than PepA-MBP. To increase the solubility of Cna peptide, RSG and GSR residues were introduced at the N- and C-terminal ends of PepA, which may have contributed to the observed increase in reconstitution (Fig. 4.11a). Closer examination of the structure of CnaB2 revealed four nearby negatively charged residues at the end of the C-terminal β-strand (Asp^{33}, Asp^{35}, Asp^{82}, and Glu^{85} shown in purple in Fig. 5.4). It was possible that the C-terminal addition of GSR to the PepA sequence in Cna peptide may have been contributing to the formation of salt bridges with Asp^{33}, Asp^{35}, Asp^{82}, or Glu^{85} (Fig. 5.4). A similar cluster of negatively charged residues was not present at the N-terminal end of the final β-strand, but the RSG residues introduced at the N-terminal end of PepA in Cna peptide may also have been contributing to increased reconstitution efficiency in an unknown way. Therefore, to further investigate the effects of RSG and GSR additions at the N- and C-terminal ends of PepA, I generated the constructs PepC, PepJ, and PepK (Table 5.3).

5.2.5.3. Rationale for PepN

Close examination of the residues in the C-terminal β-strand of CnaB2 revealed that the side chains of His^{112} and Arg^{32} were present in close proximity, and in a position that may have been favourable for electrostatic interactions (Fig. 5.4). Furthermore, a mutation at position 112 in the peptide may have been tolerated because the side chain of His^{112} was

solvent-exposed, and as a result His^{112} may not have been critical for peptide-protein reconstitution (Fig. 5.4). Therefore, I sought to investigate whether I could introduce a favourable electrostatic interaction between the positively charged Arg^{32} residue and the residue at position 112 in the peptide-protein pair. Accordingly, I mutated His^{112} in PepA to the negatively charged Glu residue to create PepN (Table 5.3).

5.2.5.4. Rationale for PepL/M/P

CnaB2 is a small domain of the larger protein FbaB where its sequence is situated in the centre of the FbaB protein (Hagan et al., 2010). It is possible that the residues that extend beyond the C-terminal end of CnaB2 (not visible in the crystal structure PDB 2X5P) interact with CnaB2 in a favourable fashion. To investigate this further, I extended the PepA peptide by 3 and 5 residues at its C-terminal end to generate PepL and PepP (Fig. 5.5 and Table 5.3). The residues that were used to extend PepA were the same are those naturally present C-terminal to the CnaB2 domain in the FbaB protein (Fig. 5.5). Similarly, I investigated increasing the length of PepA at its N-terminal end by 2 residues to assess whether a longer split peptide had favourable effects on reconstitution (Fig. 5.5 and Table 5.3).

CnaB2 Sequence: ...TKGDAHIVMVDA
FbaB Sequence: ...TKGDAHIVMVDAYKPTKGSGQVIDIEEK...

Figure 5.5. Generation of PepL, PepM, and PepP. The sequence of the C-terminal end of the CnaB2 domain that is present in the crystal structure (PDB 2X5P) is compared to the sequence that is present in FbaB including residues that extend beyond the CnaB2 domain in the complete protein (blue). The sequence of PepA is shown in red and the isopeptide bond forming D residue is underlined.

5.2.6. Reconstitution of the PepA mutant series versus SpyCatcher

5.2.6.1. PepB/D/E/F/G/H/I/O reconstitution with SpyCatcher

All of the variants of PepA were linked to MBP to aid in recombinant protein expression, and all of the constructs expressed and purified well. I then mixed the peptides individually with SpyCatcher at 10 µM of each and incubated the reactions for 30 min at 25°C. Subsequently, the reactions were analyzed and quantified using SDS-PAGE (Fig. 5.6a). Alterations in Val^{114} and Val^{116} were not well tolerated, regardless of whether a polar, negatively charged, or positively charged residue was used to replace the Val residues (Fig. 5.6b). PepB-MBP, PepF-MBP, PepG-MBP, PepH-MBP, and PepI-MBP were severely inhibited in their ability to reconstitute with SpyCatcher, each yielding less than 5% reconstitution after a 30 min reaction, whereas PepD-MBP, PepE-MBP, and PepO-MBP were slightly more active resulting in 10-15% reconstitution (Fig. 5.6b). These results demonstrated that Val^{114} and Val^{116} were required for efficient reconstitution with the protein binding partner. Individual mutation of each Val residue to polar amino acids such as Thr or Tyr led to decreased activity, while mutation of both Val residues to Thr severely reduced reconstitution efficiency (Table 5.3 and Fig. 5.6b). Furthermore, individual mutation of each Val residue to charged residues such as Lys or Glu severely reduced the activity of the peptides (Table 5.3 and Fig. 5.6b). This demonstrated that while mutations at residues Val^{114} and Val^{116} were generally not well tolerated, polar residues were better tolerated than charged residues. Therefore, the hydrophobic residues at the centre of PepA were likely critical for reconstitution and could not be altered.

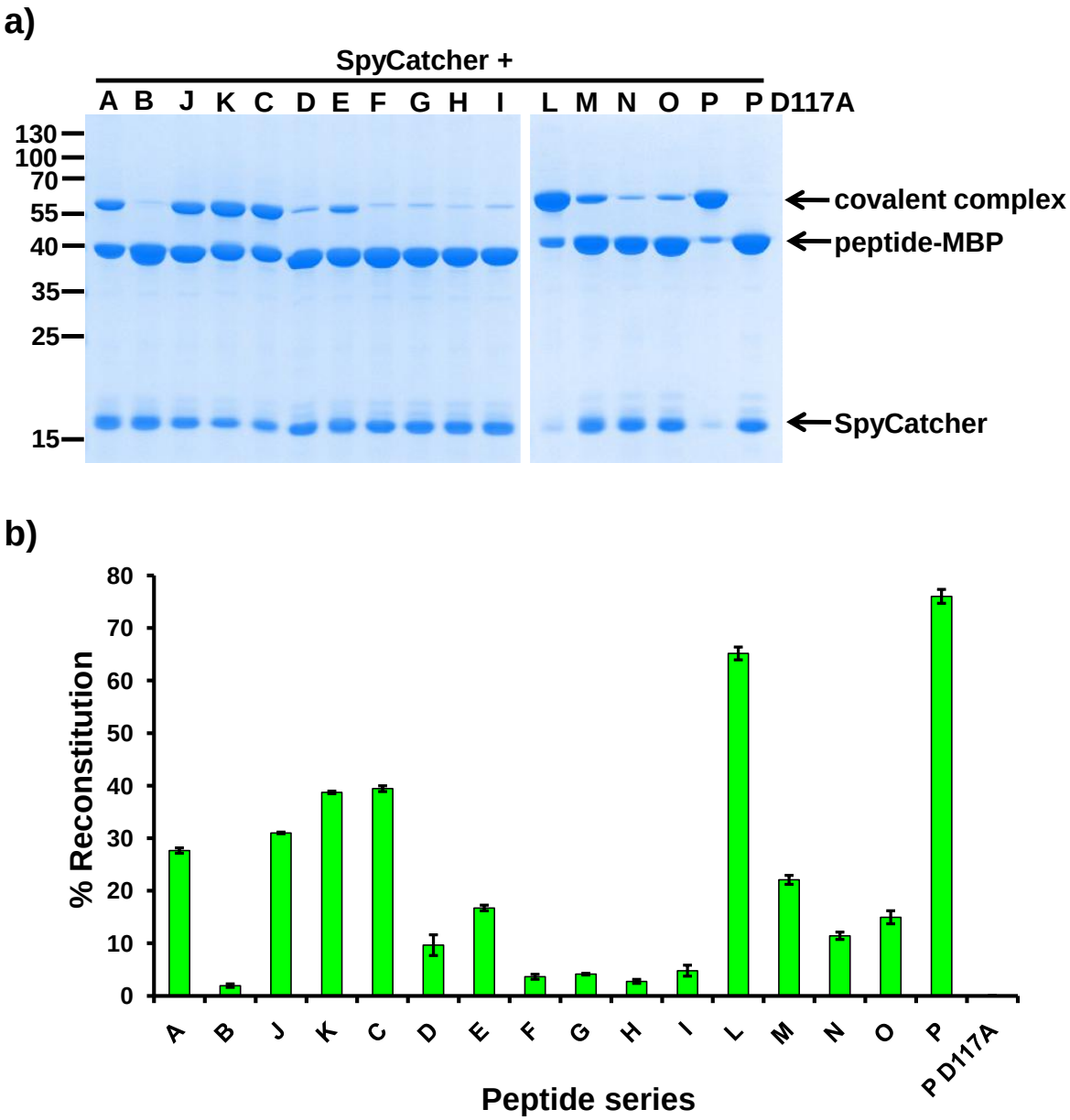


Figure 5.6. Reconstitution between SpyCatcher and PepA variants. a) SpyCatcher was mixed with the series of PepA variants linked to MBP at 10 µM of each in PBS pH 7.4 and the reactions were incubated for 30 min at 25°C. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. PepP D117A is a negative control and has the same sequence as PepP except the reactive Asp^{117} has been changed to an Ala. Samples were run on two gels, which are shown with a white bar in between. b) Quantification of the experiment in Fig. 5.6a performed in triplicate. Error bars are ± 1 S.D.

5.2.6.2. PepC/J/K reconstitution with SpyCatcher

The peptides PepC, PepJ, and PepK were made to investigate whether addition of RSG and GSR residues at the N- and C-termini of PepA would lead to more efficient reconstitution (Fig. 5.4). PepJ-MBP was not significantly different than PepA-MBP when reacted with SpyCatcher (Fig. 5.6b). However, PepK-MBP and PepC-MBP were both better at reacting with SpyCatcher than PepA-MBP, where both gave an approximately 10% higher yield in reconstitution when compared to PepA-MBP (Fig. 5.6b). PepK had the residues GSR added C-terminal to the sequence of PepA and the introduced positive charge from the Arg had a cluster of four negatively charged residues (Asp^{33}, Asp^{35}, Asp^{82}, and Glu^{85}) in a favourable proximity for it to form electrostatic interactions with (Fig. 5.4). On the other hand, since PepJ was not significantly more efficient than PepA then the increased reconstitution efficiency seen in PepC and the Cna peptide was a result of the contributions of the C-terminal GSR additions (Fig. 5.6b). Therefore, the RSG and GSR residues added at the N- and C-termini of PepA were not functioning synergistically and only the introduced C-terminal GSR enhanced reconstitution efficiency.

5.2.6.3. PepN reconstitution with SpyCatcher

His^{112} in PepA was mutated to generate PepN, with the objective of introducing a salt bridge between Glu^{112} in PepN and Arg^{32} in SpyCatcher (Fig. 5.4). However, PepN-MBP was approximately one-third less efficient at reaction with SpyCatcher than PepA-MBP (Fig. 5.6b). Although the side chain of His^{112} was solvent-exposed and possibly may not have been required for stabilizing the CnaB2 structure, mutation of His^{112} was detrimental to reconstitution showing that His^{112} was required for efficient peptide-protein interaction. These results were consistent with other peptide mutants tested (PepB/D/E/F/G/H/I/N/O), where any mutation of residues in PepA were not well tolerated (Table 5.3).

5.2.6.4. PepL/M/P reconstitution with SpyCatcher

Rather than mutating residues in PepA or extending the peptide with charged residues, I explored whether extending PepA in both the N- and C-terminal directions with residues from the wild-type FbaB sequence had any effect on reconstitution (Fig. 5.5). Accordingly, I generated PepL-MBP, PepM-MBP, and PepP-MBP and tested their ability to reconstitute with SpyCatcher. PepM-MBP (PepA peptide extended by 2 residues at its N-terminal end) was approximately 5% less efficient at reconstitution than PepA-MBP when reacted with SpyCatcher (Fig. 5.6b). However, when the PepA sequence was extended by 3 and 5 residues to make PepL and PepP, I observed a significant improvement in the ability of the peptides to reconstitute with SpyCatcher (Table 5.3). When reacted with SpyCatcher for 30 min, PepL-MBP resulted in ~65% reconstitution while PepP-MBP was even more reactive and gave a yield of ~75% (Fig. 5.6b). This enhanced reactivity may be a result of the interaction of the extra residues introduced in PepL and PepP with their surrounding residues in SpyCatcher. One contributing factor to the enhanced reactivity of PepP (where YKPTK was added to the C-terminal end of PepA) may be that the positively charged side chains of the two Lys residues may interact with the surrounding negatively charged residues Asp^{33}, Asp^{35}, Asp^{82}, and Glu^{85} (Fig. 5.4). However, other interactions such as hydrogen bonding with surrounding residues involving the newly introduced Tyr and Thr residues in PepP may also contribute to the enhanced reactivity of the peptide. As a negative control, the Asp^{117} residue in PepP that is involved in isopeptide bond formation was mutated to Ala, and as expected the resulting PepP D117A mutant did not produce a covalent complex (Fig. 5.6a).

PepP was the most efficient binding partner for SpyCatcher from the series of PepA variants explored. Therefore, PepP was renamed SpyTag and was used for all further experiments.

5.2.6.5. Reconstitution between Cna114 variants and SpyTag

During mutagenesis experiments to optimize the reconstitution efficiency of Cna114, I discovered that the I34E and M69Y mutations both enhanced the activity of Cna114 (Table 5.2). Consequently, I constructed the double mutant Cna114 I34E M69Y which was renamed SpyCatcher. To investigate the reconstitution efficiency of various Cna114 variants with the optimized SpyTag and whether the I34E and M69Y mutations acted synergistically, I tested reconstitution between Cna114, Cna114 I34E, Cna114 M69Y, and SpyCatcher with the optimize peptide SpyTag-MBP (Fig. 5.7). Constructs were mixed at 10 µM in PBS pH 7.4 and incubated at 25°C for 5 and 30 min, and then analyzed on SDS-PAGE (Fig. 5.7). Although the I34E and M69Y mutation both individually led to a significant enhancement of Cna114 reactivity, this enhancement in reactivity does not appear to be additive in the double mutant SpyCatcher.

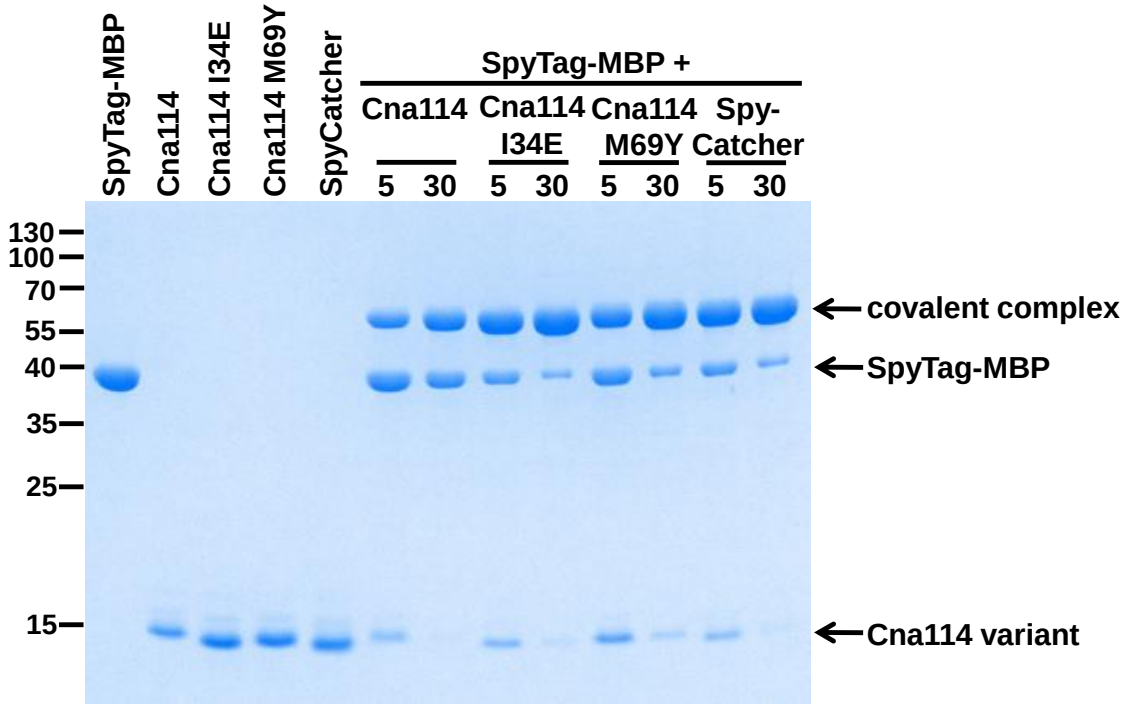


Figure 5.7. Reconstitution between SpyTag-MBP and Cna114 variants. Proteins were each mixed at 10 µM in PBS pH 7.4 and incubated at 25°C for 5 and 30 min. Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining.

5.3. Summary

To improve the reconstitution efficiency of the split CnaB2 peptide-protein pair, I used phage display and rational modifications to optimize the peptide and protein sequences for reconstitution. I used phage display to try to isolate an efficient covalently binding peptide for Cna110 and Cna114 from a large random peptide library. However, I was unable to find any covalently binding peptides. Based on mutagenesis experiments on PepA, it was clear that alterations of residues in the C-terminal β-sheet of CnaB2 were not well tolerated (Table 5.3). Therefore, it was not surprising that covalently binding peptides were not isolated from a random peptide library, as the peptides in the phage display library tested did not cover all of the sequence diversity possible in a 12 residue peptide (New England Biolabs, 2009), and the probability of phage display peptides containing the PepA sequence was very low. A better approach for phage display experiments in future studies would be to develop a peptide library that contains the core PepA sequence required for efficient reconstitution, but explores sequence diversity in residues flanking the PepA sequence.

Consequently, I tried a rational modification approach where I attempted to optimize both Cna114 and PepA by introducing a series of mutations to enhance the binding interactions between the peptide-protein pair. This approach was fruitful and led to the isolation of a much improved binding partner in SpyCatcher, which incorporated the I34E and M69Y mutations to the Cna114 platform. To optimize PepA, I investigated a series of mutations in the peptide sequence targeted to increase peptide solubility and introduce novel electrostatic interactions between the peptide-protein pair. Although introduction of a positively charged residue at the C-terminal end of the peptide resulted in more efficient reconstitution, alterations to the core peptide sequence were not well tolerated. The core peptide sequence which constituted the C-terminal β-strand in CnaB2 was likely necessary during peptide-protein reconstitution for establishing the core hydrophobic environment and the correct positioning of the isopeptide bonding triad, which is required for spontaneous

isopeptide bond formation (Kang and Baker, 2011; Hu et al., 2010). Alterations in the PepA sequence likely disrupted the hydrophobic environment of the reconstituted protein or oriented the isopeptide bonding triad in a conformation less efficient for reconstitution. However, I discovered significantly improved reconstitution efficiency in the peptides by extending the C-terminal sequence of PepA by 3 and 5 residues to make PepL, and the most efficiently reacting peptide SpyTag. The optimized peptide-protein pair of SpyTag and SpyCatcher were significantly improved from the starting platforms of PepA and Cna114, and their reaction yielded ~75% reconstitution in 30 min. To further understand the nature of this interaction, I proceeded to characterize the optimized SpyTag:SpyCatcher technology.

Chapter 6: Characterization of the SpyTag:SpyCatcher technology

Chapter 6: Characterization of the SpyTag:SpyCatcher technology

CnaB2 is a small domain of the larger FbaB protein, a fibronectin binding protein from *S. pyogenes* (Hagan et al., 2010). I dissected CnaB2 to construct an 8 residue peptide and a small binding partner that were able to reconstitute via intermolecular isopeptide bond formation. The split CnaB2 constructs were more efficient at reconstitution than their Spy0128 counterparts. However, in an attempt to further improve the efficiency of reconstitution of the split CnaB2 constructs, I constructed and analyzed a series of variants of both PepA and Cna114. These analyses led to the generation of SpyTag and SpyCatcher, a peptide-protein pair that were significantly more efficient at reconstitution than PepA and Cna114. Herein, I characterized the covalent reconstitution of SpyTag and SpyCatcher under a variety of conditions relevant for biological experiments.

6.1. The SpyTag:SpyCatcher technology

Earlier, I had split the CnaB2 protein domain to create the 8 residue peptide PepA (CnaB2 residues 111-118), and the binding partner Cna114 (CnaB2 residues 1-113). I then proceeded to optimize the efficiency of reconstitution between this peptide-protein pair by extending PepA by 5 residues from the FbaB sequence at its C-terminal end to create SpyTag, and I introduced the two mutations I34E and M69Y in Cna114 to develop SpyCatcher (Fig. 6.1a). When SpyTag-MBP and SpyCatcher were mixed at 10 μM each in PBS pH 7.4 and incubated for 3 hr at 25°C followed by SDS-PAGE analysis, a new product formed at ~60 kDa corresponding to a covalently complexed SpyTag-MBP:SpyCatcher pair (Fig. 6.1b). However, when the catalytic Glu^{77} in SpyCatcher was mutated to a Gln, SpyCatcher E77Q was no longer able to form a covalent complex with SpyTag (Fig. 6.1b). Similarly, when the reactive Asp^{117} in SpyTag was mutated to Ala, SpyTag-MBP D117A no

longer formed a covalent complex with SpyCatcher (Fig. 6.1b). Furthermore, reconstitution between SpyTag-MBP and SpyCatcher yielded only one observable covalent complex and no side-products were observed as with the isopeptag:pilin-C system (compare Fig. 6.1b with Fig. 3.9b).

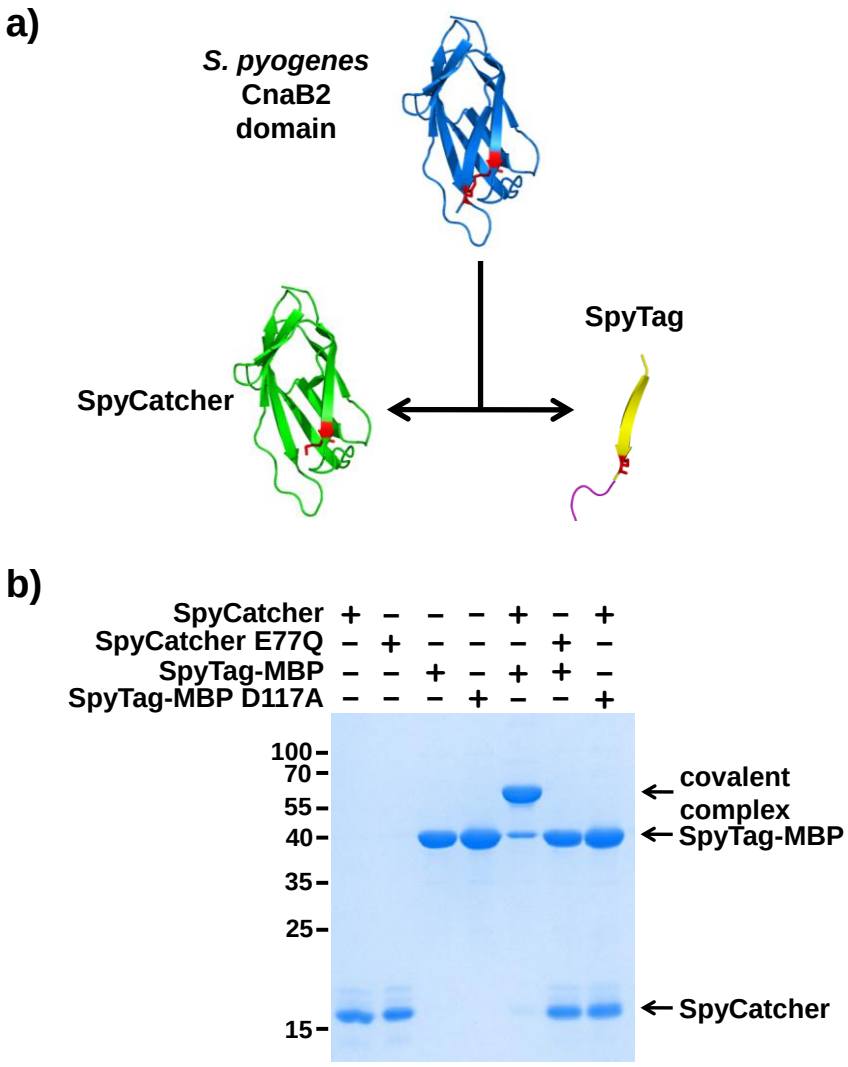


Figure 6.1. Covalent reconstitution between SpyTag-MBP and SpyCatcher. a) Cartoon illustrating the construction of SpyTag and SpyCatcher. CnaB2 (blue) (PDB 2X5P) was split to form Cna114 (CnaB2 residues 1-113), which was then mutated to form SpyCatcher (green) (Cna114 I34E M69Y). SpyTag (yellow) was formed by extending the final 8 residues of CnaB2 with 5 residues from the FbaB sequence (purple). Isopeptide bond forming residues are shown in red. b) SpyTag-MBP and SpyCatcher were able to covalently reconstitute. Proteins were mixed at 10 µM in PBS pH 7.4 and incubated for 3 hr at 25°C.

Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. The catalytic Glu^{77} was mutated to Gln in the unreactive control SpyCatcher E77Q. The reactive Asp^{117} was mutated in Ala in the unreactive control SpyTag-MBP D117A.

6.2. Mass spectrometry of the SpyCatcher and Cna peptide reaction

To demonstrate the covalent nature of the interactions between SpyTag and SpyCatcher, I attempted to analyze the reaction between SpyTag-MBP and SpyCatcher using mass spectrometry. However, I was unable to obtain a result with the required accuracy to verify the formation of an isopeptide bond between the two proteins. This was likely a result of the instrument being unable to accurately measure the mass of large protein complexes. Therefore, to illustrate isopeptide bond formation between SpyCatcher and a covalently binding peptide, I analyzed reconstitution between SpyCatcher and Cna peptide (sequence: RSGAHIVMVDAGSR). I mixed SpyCatcher with a 2-fold excess of Cna peptide in PBS pH 7.4 and incubated the mixture for 16 hr at 4°C to minimize potential sample degradation by proteases. Subsequently, I dialyzed the reaction and analyzed it using mass spectrometry (Fig. 6.2). SpyCatcher had an expected mass of 15,240.5 Da and Cna peptide had an expected mass of 1,455.7 Da. If SpyCatcher and Cna peptide reacted to form a covalent complex via an isopeptide bond, then a water molecule would be released and the expected mass of the covalent complex would be 16,678.2 Da (15,240.5 Da + 1,455.7 Da – 18 Da). The reaction between SpyCatcher and Cna peptide yielded a product with a mass of 16,678.3 Da (Fig. 6.2), which was in close agreement with the expected mass of a covalently complex SpyCatcher:Cna peptide pair.

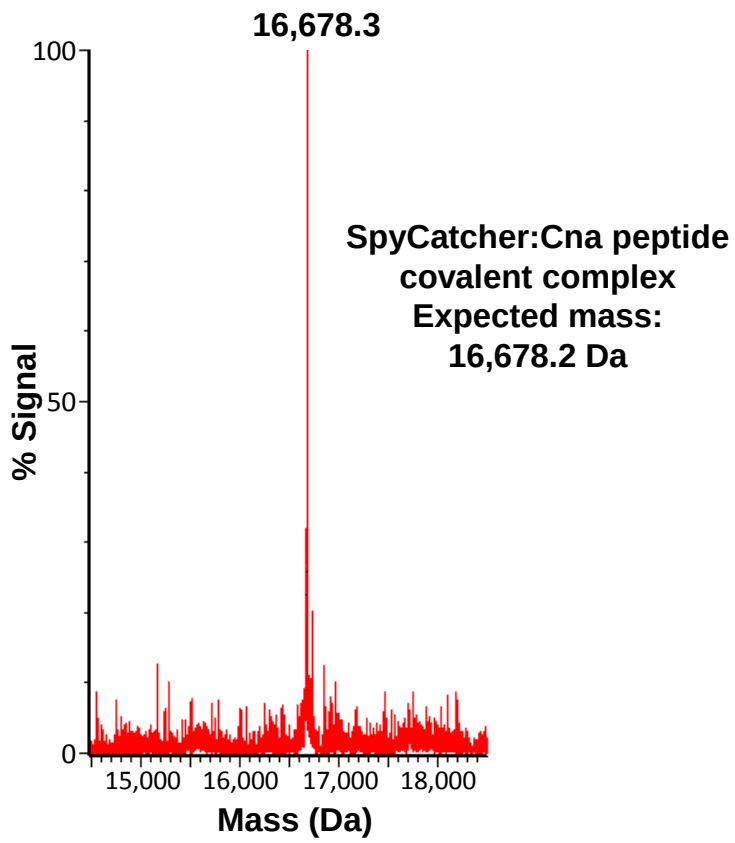


Figure 6.2. Mass spectrometry of the SpyCatcher and Cna peptide reaction. SpyCatcher at 100 µM was mixed with Cna peptide at 200 µM in PBS pH 7.4 and incubated at 4°C for 16 hr. The reaction was then dialyzed 3 times against MilliQ water and analyzed by mass spectrometry. Based on their predicted sequences, SpyCatcher had an expected mass of 15,240.5 Da and Cna peptide has an expected mass of 1,455.7 Da. If they react via isopeptide bond formation, then a H_2O molecule would be released and the covalently complex product would have an expected mass of 16,678.2 Da (15240.5 Da + 1455.7 Da – 18.0 Da).

6.3. Speed of reconstitution between SpyTag-MBP and SpyCatcher

To investigate the speed of reconstitution, I performed a time course experiment for the reaction between SpyTag-MBP and SpyCatcher. I mixed both proteins at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated the reaction at 25°C. Subsequently, samples were removed at various time points and analyzed on SDS-PAGE (Fig. 6.3a), from which the reactions were quantified. Speed of reconstitution was assessed by measuring the depletion of SpyCatcher versus time (Fig. 6.3b). The reaction proceeded rapidly over the first

few minutes where over 40% of SpyCatcher was depleted after 1 min of incubation with SpyTag-MBP, and over 90% had reacted after 10 min (Fig. 6.3b). The reaction then started to plateau and reached 98% SpyCatcher depletion after 30 min, and after 120 min SpyCatcher can no longer be detected (Fig. 6.3b). Overall, the speed of reaction between SpyTag-MBP and SpyCatcher was significantly improved when compared to the original split CnaB2 peptide-protein pair (compare Fig. 6.3b and Fig. 4.9b).

Using the data from the time course experiment that I performed to determine the speed of reconstitution, I proceeded to calculate the rate constant for the reaction. Reconstitution measured as depletion of SpyCatcher proceeded linearly for the first 5 min of reaction after which it started to slow down (Fig. 6.3b). To calculate the rate constant for the reaction, I constructed a plot of 1/[SpyCatcher] versus time over the first 5 min of reaction for which it proceeded linearly (Fig. 6.3c) (Price et al., 2001). This produced a linear plot which signified an overall second order reaction (Fig. 6.3c) (Price et al., 2001). Based on the gradient of the trendline (Fig. 6.3c), the rate constant (k) of the SpyCatcher and SpyTag-MBP reaction was 1.4×10^3 $M^{-1}s^{-1}$, which was used to calculate a reaction half-time of 74 seconds ($t_{1/2}=1/k[\text{SpyCatcher}]_0$) (Fig. 6.3c) (Price et al., 2001).

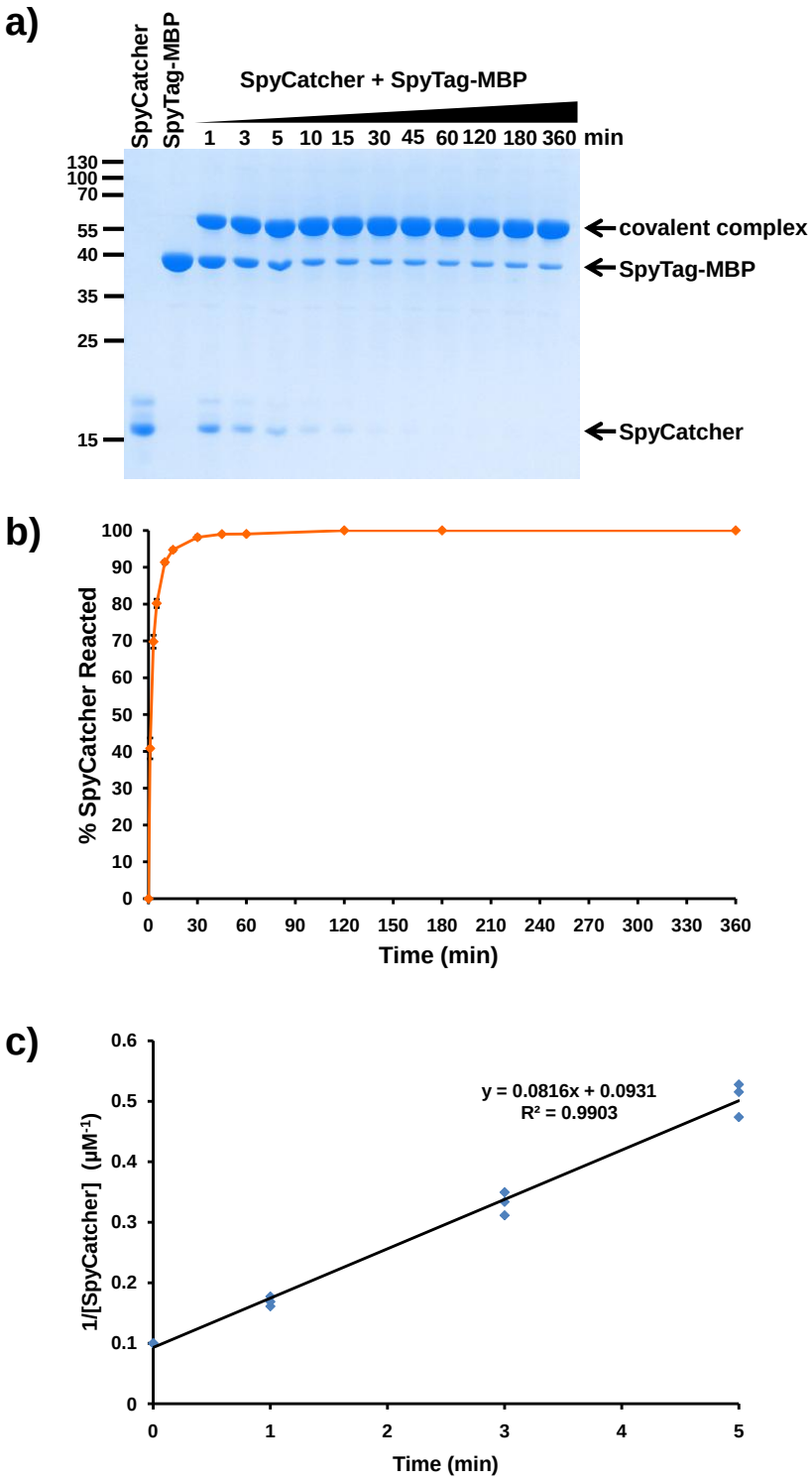


Figure 6.3. Speed of reconstitution between SpyTag-MBP and SpyCatcher. a) Proteins were mixed at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25°C for various time points. Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. b) Reactions from the experiments in Fig. 6.3a were quantified by measuring the depletion of the band

corresponding to SpyCatcher. The results of triplicate reactions are shown ±1 S.D. Error bars are too small to be seen at some time points. c) Calculation of the rate constant for the SpyTag-MBP and SpyCatcher reaction. Using the data points in triplicate from the experiment shown in Fig. 6.3a, a plot of 1/[SpyCatcher] versus time (over the first 5 min of reaction when it proceeded linearly) was made. Based on the gradient of the trendline, the reaction had a rate constant (k) of 1.4×10^3 $M^{-1}s^{-1}$.

6.4. SpyTag-MBP and SpyCatcher reconstitution at varying ratios

Thus far, covalent reconstitution between SpyTag-MBP and SpyCatcher had only been investigated when both partners were present at equal molar concentrations. To investigate the effect on reconstitution efficiency of having one partner in excess, I tested the SpyTag-MBP and SpyCatcher reaction at varying ratios. One partner was maintained at a constant concentration of 10 µM while the other partner was mixed at concentrations of 0, 1, 2.5, 5, 10 and 20 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and the reactions were then incubated at 25°C for 3 hr. Subsequently, the samples were analyzed on SDS-PAGE, where the reactions were quantified by measuring the depletion of the partner maintained at a constant concentration of 10 µM (Fig. 6.4). Both SpyTag-MBP and SpyCatcher produced similar reaction profiles when maintained at 10 µM, where neither partner required the other to be present in excess to achieve high levels of reconstitution (Fig. 6.4). This was in contrast to the isopeptag:pilin-C pair where excess amounts of each binding partner were required to achieve high levels of reconstitution (Fig. 3.12). Interestingly, under the experimental conditions tested over 90% of SpyCatcher was depleted when mixed with SpyTag-MBP at a ratio of 1:1, however when SpyTag-MBP was present at a 2-fold excess in concentration all of SpyCatcher was depleted (Fig. 6.4). On the other hand, when SpyTag-MBP was present at a constant concentration of 10 µM, only 75% of SpyTag-MBP was depleted when mixed with SpyCatcher at a ratio of 1:1 and complete depletion of SpyTag-MBP was not observed when SpyCatcher was present in 2-fold excess (Fig. 6.4). It would be expected that when both partners are present at equal molar concentration then the same level of reconstitution

should be expected regardless of which protein is being maintained at 10 µM. However, the difference observed was likely a result of the method by which the reactions were quantified, where I measured the decrease in signal intensity of either the SpyTag-MBP or SpyCatcher band. Since SpyTag-MBP was much larger than SpyCatcher, on a SDS-PAGE gel SpyTag-MBP could be better detected at lower concentrations.

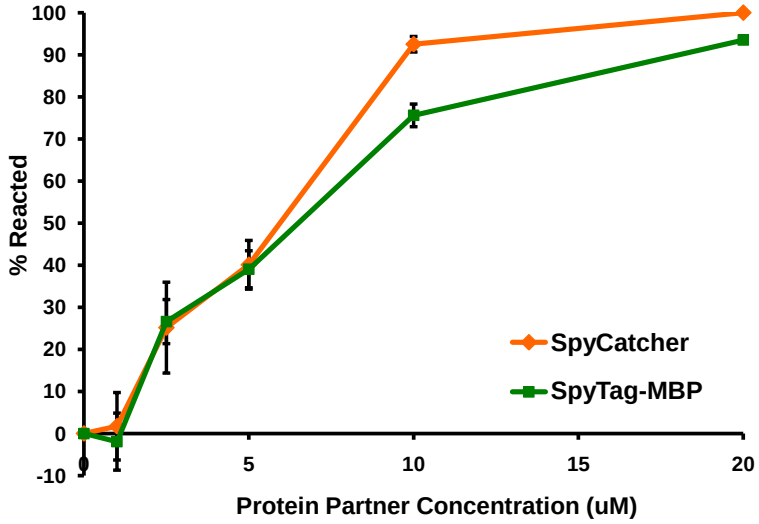


Figure 6.4. SpyTag-MBP and SpyCatcher reconstitution at varying protein ratios. Either SpyCatcher (orange) or SpyTag-MBP (green) was maintained at a constant concentration of 10 µM while the other partner was mixed at concentrations of 0, 1, 2.5, 5, 10 and 20 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and the reactions were then incubated at 25°C for 3 hr. Subsequently, reactions were heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. Results were quantified by measuring the depletion of the partner maintained at 10 µM and the mean of triplicate reactions is shown ±1 S.D.

6.5. Concentration-dependence of SpyTag-MBP and SpyCatcher reconstitution

In many biological assays proteins and hence their corresponding peptide fusion tags are present at concentrations below 10 µM. Therefore, it was important to determine whether SpyTag and SpyCatcher were able to efficiently reconstitute at lower concentrations. Accordingly, I assessed the concentration-dependence of SpyTag-MBP and SpyCatcher reconstitution by mixing each protein at equal molar concentrations of 1, 5, or 10 µM in 40

mM Na_2HPO_4 – 20 mM citric acid pH 7.0, and incubating the reactions at 25°C for various time points. I then analyzed and quantified the reactions using SDS-PAGE (Fig. 6.5). As expected, reconstitution was less efficient with lower concentrations at earlier time points, however after a 60 min incubation period all of the concentrations tested yielded similar levels of covalent product (Fig. 6.5). This was in contrast to the isopeptag:pilin-C pair where reconstitution was significantly less efficient at lower concentrations (Fig. 3.13). Therefore, SpyTag and SpyCatcher were able to covalently react efficiently even at lower protein concentrations.

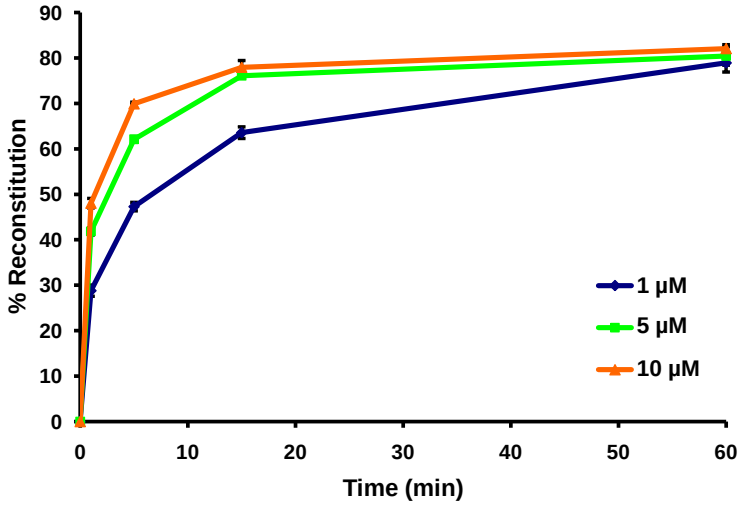


Figure 6.5. Concentration-dependence of SpyTag-MBP and SpyCatcher reconstitution. SpyTag-MBP and SpyCatcher were mixed at equal concentrations of either 1, 5, or 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 25°C for various time points. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. Reactions were quantified by measuring product formation and the mean of triplicate reactions are shown ±1 S.D. Some error bars are too small to be visible on the graph.

6.6. Temperature-dependence of SpyTag-MBP and SpyCatcher reconstitution

It would be ideal for a peptide fusion tag to efficiently function at the range of temperatures from 4 to 37°C. When peptide fusion tags are used for protein purifications

lower temperatures are ideal to minimize protein degradation, while activity under physiological temperatures would be desired for their use in *in vivo* studies. Therefore, I tested the ability of SpyTag-MBP and SpyCatcher to reconstitute at various temperatures by mixing each protein at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubating the reactions at 4, 25, and 37°C for 1 and 10 min. Subsequently, I analyzed and quantified the reactions on SDS-PAGE (Fig. 6.6). Surprisingly, SpyTag-MBP and SpyCatcher were able to covalently reconstitute efficiently at all incubation temperatures tested (Fig. 6.6). It was interesting to observe that reconstitution was slightly less efficient at 4°C as compared to 25 and 37°C, where approximately 20% less product formation occurred after a 1 min reaction and 15% less product was observed after a 10 min incubation (Fig. 6.6). This was in contrast to the isopeptag:pilin-C pair where reconstitution was slightly more efficient at 4 and 25°C as compared to 37°C (Fig. 3.14).

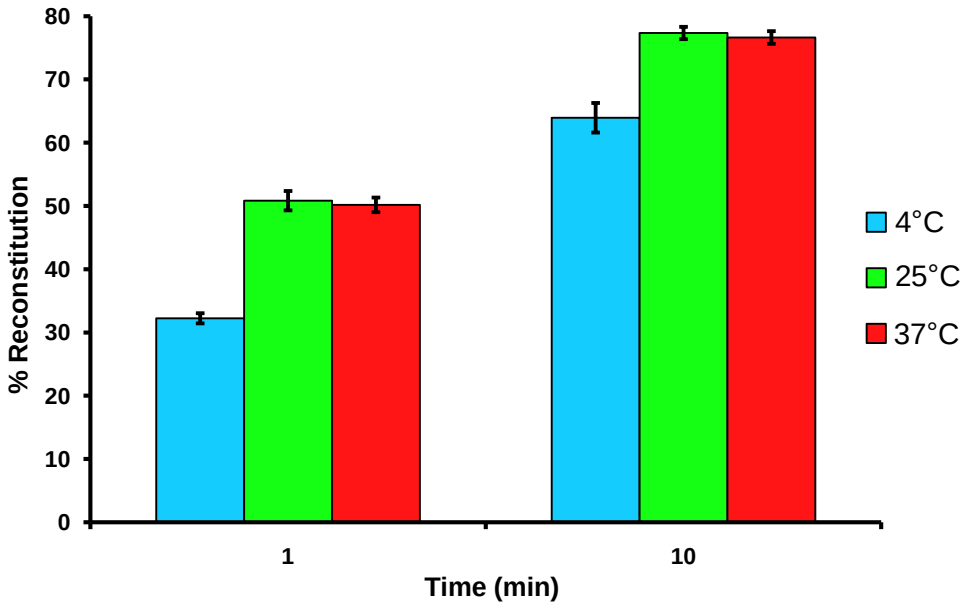


Figure 6.6. Temperature-dependence of SpyTag-MBP and SpyCatcher covalent reconstitution. SpyTag-MBP and SpyCatcher were mixed at 10 µM each in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated at 4, 25, and 37°C for 1 and 10 min. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14%

SDS-PAGE with Coomassie staining. Reactions were quantified by measuring product formation and the mean of triplicate reactions are shown ±1 S.D.

6.7. pH-dependence of SpyTag-MBP and SpyCatcher reconstitution

To investigate the dependency of covalent reconstitution between SpyTag and SpyCatcher on the pH of the environment, I tested their activity at a range of pH values. I mixed SpyTag-MBP and SpyCatcher each at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid buffer with pH values adjusted to 5, 6, 7, and 8. I then incubated the reactions at 25°C for 1 and 10 min, and then analyzed and quantified the reactions on SDS-PAGE (Fig. 6.7). Covalent reconstitution between SpyTag-MBP and SpyCatcher was robust and occurred at all of the pH values tested, where at least 50% covalent complex formation was observed after a 10 min incubation period (Fig. 6.7). Interestingly, at the early time point of 1 min reconstitution was most efficient at lower pH values of 5 and 6 where approximately 60% product yield was obtained, and product yield was reduced by 10% at pH 7 and 35% at pH 8 (Fig. 6.7). However, at the later time point of 10 min reconstitution yields were similar at pH values of 5, 6, and 7, whereas it was still lower at pH 8 by 20% (Fig. 6.7). This was in contrast to the isopeptag:pilin-C pair where reconstitution was most efficient at pH 7 and 8 and least efficient at pH 5 (Fig. 3.15).

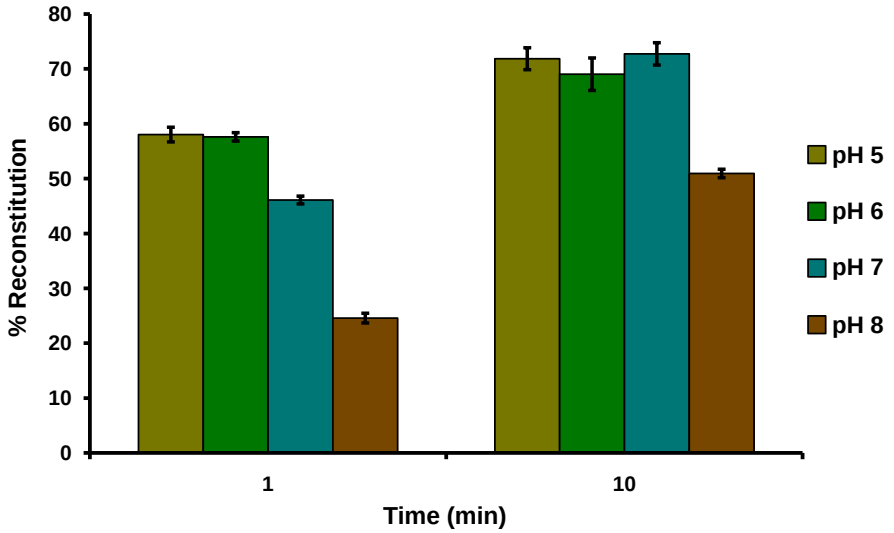


Figure 6.7. SpyTag-MBP and SpyCatcher reconstitution proceeded at a range of pH values. SpyTag-MBP and SpyCatcher were mixed at 10 µM each in 40 mM Na_2HPO_4 – 20 mM citric acid buffer adjusted to pH 5.0, 6.0, 7.0, and 8.0, then incubated at 25°C for 1 and 10 min. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. Reactions were quantified by measuring product formation and the mean of triplicate reactions are shown ±1 S.D.

6.8. SpyTag-MBP and SpyCatcher reconstitution in commonly used buffers

To assess whether the efficiency of reconstitution was dependent on buffer composition, I investigated the ability of SpyTag-MBP and SpyCatcher to reconstitute in a range of commonly used biological buffers. I mixed SpyTag-MBP and SpyCatcher at 10 µM each in PBS pH 7.4, 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0, 50 mM Tris pH 7.0, and 50 mM HEPES pH 7.0, and incubated the reactions at 25°C for 1 and 10 min. I then analyzed the reactions on SDS-PAGE, and quantified the results (Fig. 6.8). SpyTag-MBP and SpyCatcher were able to covalently reconstitute in all of the buffers tested (Fig. 6.8). However, reconstitution was most efficient in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and 50 mM HEPES pH 7.0 where over 70% product yield was observed after a 10 min reaction (Fig. 6.8). Although, reconstitution efficiency was slightly reduced in 50 mM Tris pH 7.0, SpyTag-MBP and SpyCatcher reconstitution was least efficient in PBS pH 7.4, where after a

10 min reaction there was 10% less product formation compared to reaction in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 (Fig. 6.8).

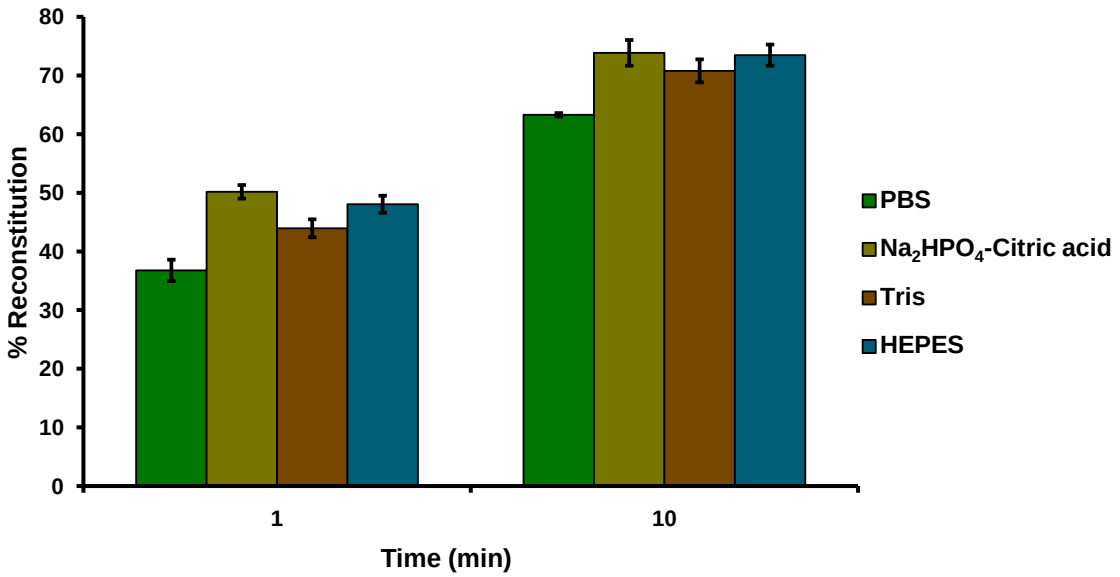


Figure 6.8. SpyTag-MBP and SpyCatcher reconstitution in a range of biological buffers. SpyTag-MBP and SpyCatcher were mixed at 10 µM each in PBS pH 7.4, 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0, 50 mM Tris pH 7.0, and 50 mM HEPES pH 7.0, and incubated at 25°C for 1 and 10 min. Samples were then taken and heated to 95°C for 7 min in SDS loading buffer and run on 14% SDS-PAGE with Coomassie staining. Reactions were quantified by measuring product formation and the mean of triplicate reactions are shown ±1 S.D.

6.9. SpyTag-MBP and SpyCatcher reconstitution in the presence of detergent

Detergents are often used in biochemical experiments to disrupt biological membranes or non-covalent protein-protein interactions. To assess whether SpyTag could be used as a peptide fusion tag under these conditions, I investigated the ability of SpyTag-MBP and SpyCatcher to covalently reconstitute in the presence of commonly used detergents. Since SpyCatcher was a split protein fragment present in a partially disordered state, it was possible that detergents may have interacted with the hydrophobic core of the protein and

interfered with reconstitution (Voet and Voet, 2004). To test this, I mixed SpyTag-MBP and SpyCatcher each at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 containing either no detergent, 1% Triton-X 100, 1% Tween20, or 0.5% Nonidet P40 and incubated the reactions for 3 hr at 25°C. Subsequently, I analyzed the reactions on SDS-PAGE, and quantified the results (Fig. 6.9). Surprisingly, the presence of detergent did not interfere with reconstitution of SpyTag-MBP and SpyCatcher, where similar levels of covalently complexed SpyTag-MBP:SpyCatcher were observed in the absence of detergent or in the presence of 1% Triton-X 100, 1% Tween20, or 0.5% NP40 (Fig. 6.9). Therefore, the SpyTag:SpyCatcher technology can be used to link two polypeptide sequences in the present of commonly used detergents.

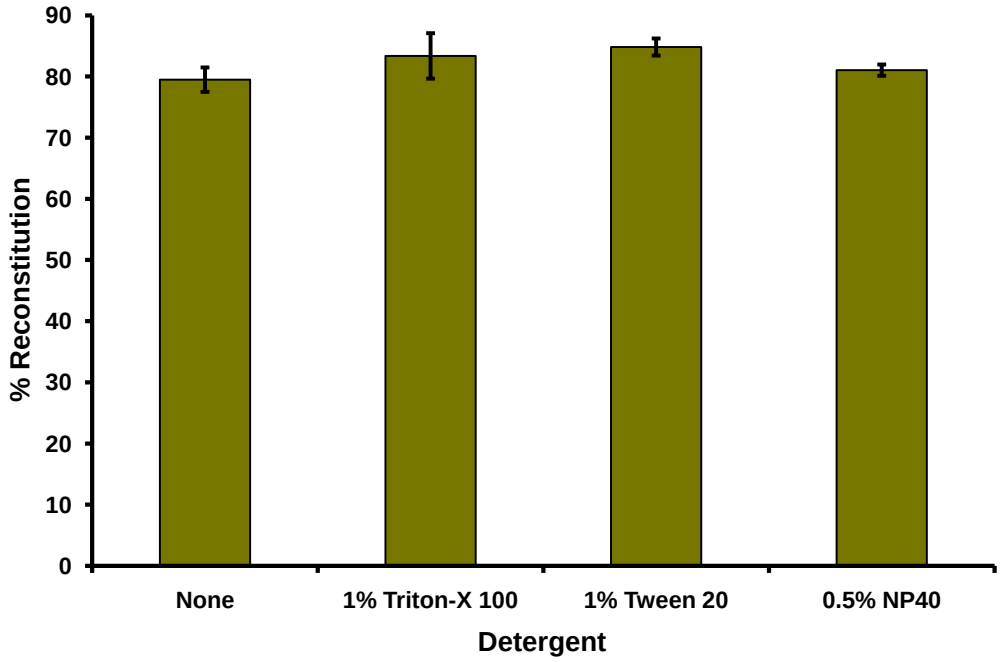


Figure 6.9. SpyTag-MBP and SpyCatcher reconstituted in the presence of commonly used detergents. SpyTag-MBP and SpyCatcher were mixed at 10 µM each in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 containing either no detergent or containing 1% Triton-X 100, 1% Tween20, or 0.5% NP40. The reactions were then incubated at 25°C for 3 hr. Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. Reactions were quantified by measuring product formation and the mean of triplicate reactions are shown ±1 S.D.

6.10. Reversibility test of the SpyTag-MBP:SpyCatcher covalent reaction

The Spy0128 and ACE19 scaffolds that I split to construct covalent peptide-protein pairs both contained intramolecular isopeptide bonds that formed between Lys and Asn side chains. When Lys and Asn side chains react to form an isopeptide bond a NH_3 molecule is released in to the aqueous environment. Hence, NH_3 would be present at negligible concentrations and it would be unlikely to contribute to reversal of a Lys-Asn isopeptide bond. However, in CnaB2 the isopeptide bond forms between Lys and Asp side chains leading to the release of H_2O, which is present in large excess in the surrounding aqueous environment (~55 M). Therefore, it may have been possible for the intermolecular isopeptide bond that formed between Lys and Asp side chains in the SpyTag-MBP:SpyCatcher covalent complex to be reversible via hydrolysis (Fig. 6.10a).

To test whether the SpyTag-MBP:SpyCatcher covalent complex was reversible, I incubated SpyTag-MBP and SpyCatcher each at 10 µM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 for 3 hr at 25°C. I then added 20-fold excess of Cna peptide and incubated the mixture for 16 hr at 25°C. Subsequently, I heated all of the reactions to 95°C for 7 min in SDS loading buffer to break all non-covalent interactions and analyzed the reactions on SDS-PAGE (Fig. 6.10b). If the isopeptide linkage between SpyTag-MBP and SpyCatcher was reversible, then it would be expected that upon incubation with excess Cna peptide some of the SpyTag-MBP:SpyCatcher covalent complexes would be replaced by Cna peptide:SpyCatcher covalent complexes. On SDS-PAGE this would appear as a reduction in the intensity of the band for the SpyTag-MBP:SpyCatcher covalent complex and the appearance of a new band for the Cna peptide:SpyCatcher covalent complex. When Cna peptide was incubated with SpyCatcher a new band formed corresponding to a Cna peptide:SpyCatcher covalent complex (Fig. 6.10b). However, when SpyTag-MBP and SpyCatcher were allowed to react for 3 hr and then Cna peptide was added to the mixture for 16 hr, there was no reduction in the intensity of the band for the SpyTag-MBP:SpyCatcher

covalent complex and a band for the Cna peptide:SpyCatcher covalent complex was not present (Fig. 6.10b). As negative controls, the unreactive SpyCatcher E77Q was incubated with either SpyTag-MBP or Cna peptide and neither reaction led to covalent complex formation (Fig. 6.10b). Therefore, the intermolecular isopeptide bond that formed between SpyTag-MBP and SpyCatcher was not reversible under these conditions.

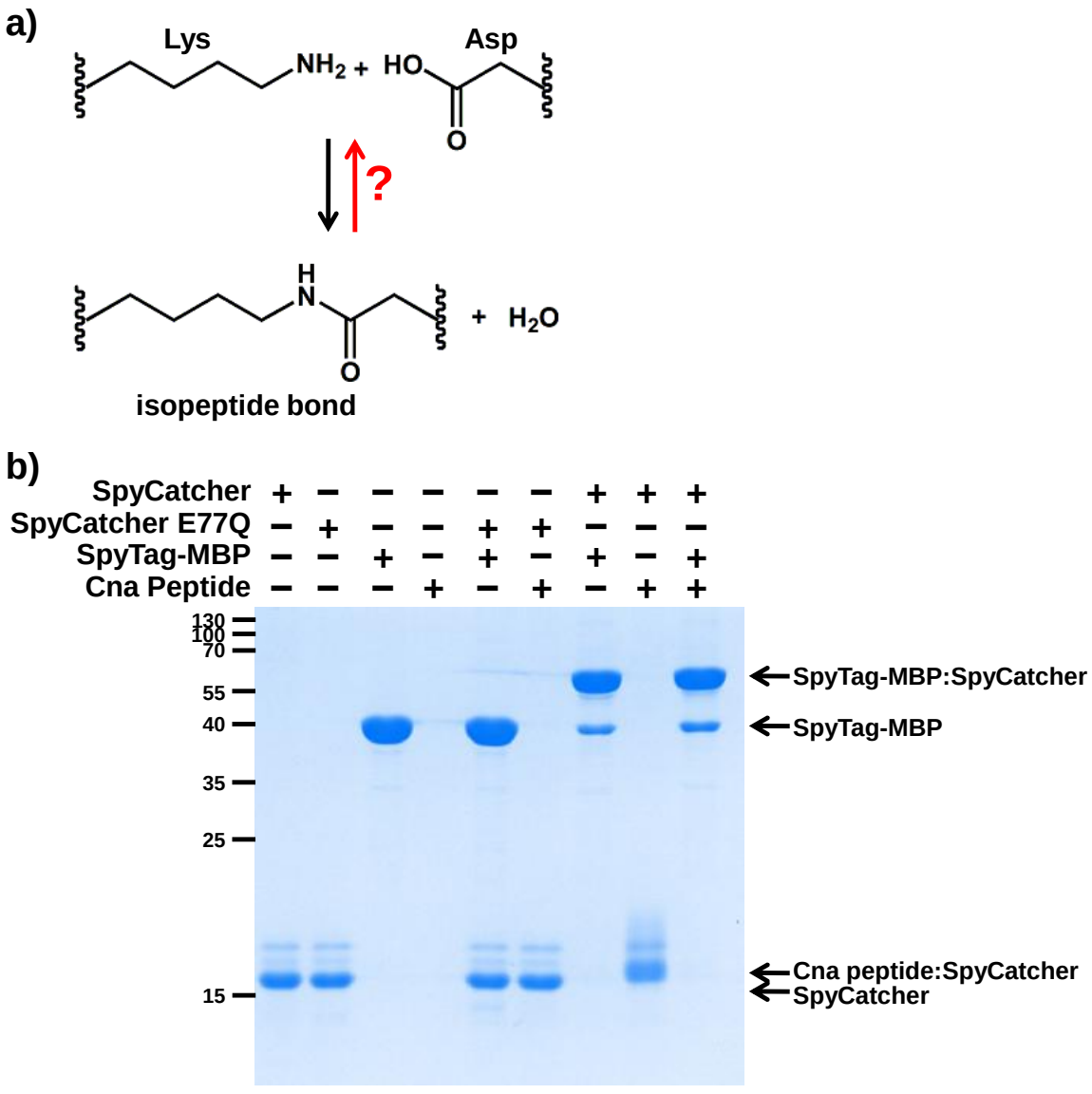


Figure 6.10. SpyTag-MBP and SpyCatcher reconstitution was not reversible. a) Isopeptide bond formation between Lys and Asp side chains in CnaB2 releases a H_2O molecule. The resulting isopeptide bond may be susceptible to hydrolysis. b) A test of the

reversibility of the SpyTag-MBP:SpyCatcher covalent complex. Proteins were mixed at 10 μM and peptides were mixed at 200 μM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated for 3 hr at 25°C. Subsequently, to one of the SpyTag-MBP and SpyCatcher reactions Cna peptide was added at 200 μM and all reaction were incubated for a further 16 hr at 25°C. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 18% SDS-PAGE with Coomassie staining. The catalytic Glu^{77} in SpyCatcher was mutated to Gln to make the unreactive control SpyCatcher E77Q.

6.11. Cross-reactivity of the SpyTag:SpyCatcher and isopeptag:pilin-C technologies

CnaB2 and Spy0128 are both outer-membrane proteins of the Gram-positive bacterial pathogen *S. pyogenes*, and both proteins utilize intramolecular isopeptide bonds to stabilize their structures (Kang et al., 2007; Hagan et al., 2010). Furthermore, both CnaB2 and Spy0128 are composed of all β-strands with their intramolecular isopeptide bonds, which are surrounded by a cluster of hydrophobic residues, connecting the first and last β-strand in their domains (Kang et al., 2007; Hagan et al., 2010). Therefore, I sought to investigate whether the SpyTag:SpyCatcher and isopeptag:pilin-C pairs that were generated from the proteins were able to cross-react. Inspection of the sequences of the two peptides revealed that the four residues N-terminal to the reactive Asp (SpyTag) or Asn (isopeptag) shared similar chemical properties (Fig. 6.11a in blue). Either hydrophobic residues were present in the same positions in both peptides or the two Val residues in SpyTag were replaced by Thr residues in isopeptag, which was a conservative change (Fig. 6.11a in blue). I mixed SpyTag-MBP with pilin-C, isopeptag-MBP with SpyCatcher, and the appropriate control reactions each at 10 μM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated the reactions for 24 hr at 25°C. Subsequently, the reactions were analyzed on SDS-PAGE (Fig. 6.11b). Surprisingly, isopeptag-MBP was able to covalently react with SpyCatcher although with a low product yield, however SpyTag-MBP was not able to covalently complex with pilin-C (Fig. 6.11b). Even though the reactive residue in isopeptag was Asn while in SpyTag it was Asp (Fig. 6.11a), isopeptag still showed compatibility with SpyCatcher demonstrating that the two technologies were able to cross-react (Fig. 6.11b). The observed cross-reactivity was likely

because the cluster of 4 similar residues N-terminal to the reactive Asn allowed isopeptag to bind to SpyCatcher (Fig. 6.11a), then position the reactive Asn in the correct orientation and establish the hydrophobic core necessary for spontaneous isopeptide bond formation (Kang and Baker, 2011). Interestingly, covalent reconstitution between isopeptag-MBP and SpyCatcher did not result in the formation of side-products, unlike reconstitution between isopeptag-MBP and pilin-C (Fig. 6.11b). This further demonstrated that the side-products that formed as a result of the isopeptag-MBP and pilin-C reaction was likely due to covalent dimerization of two pilin-C molecules.

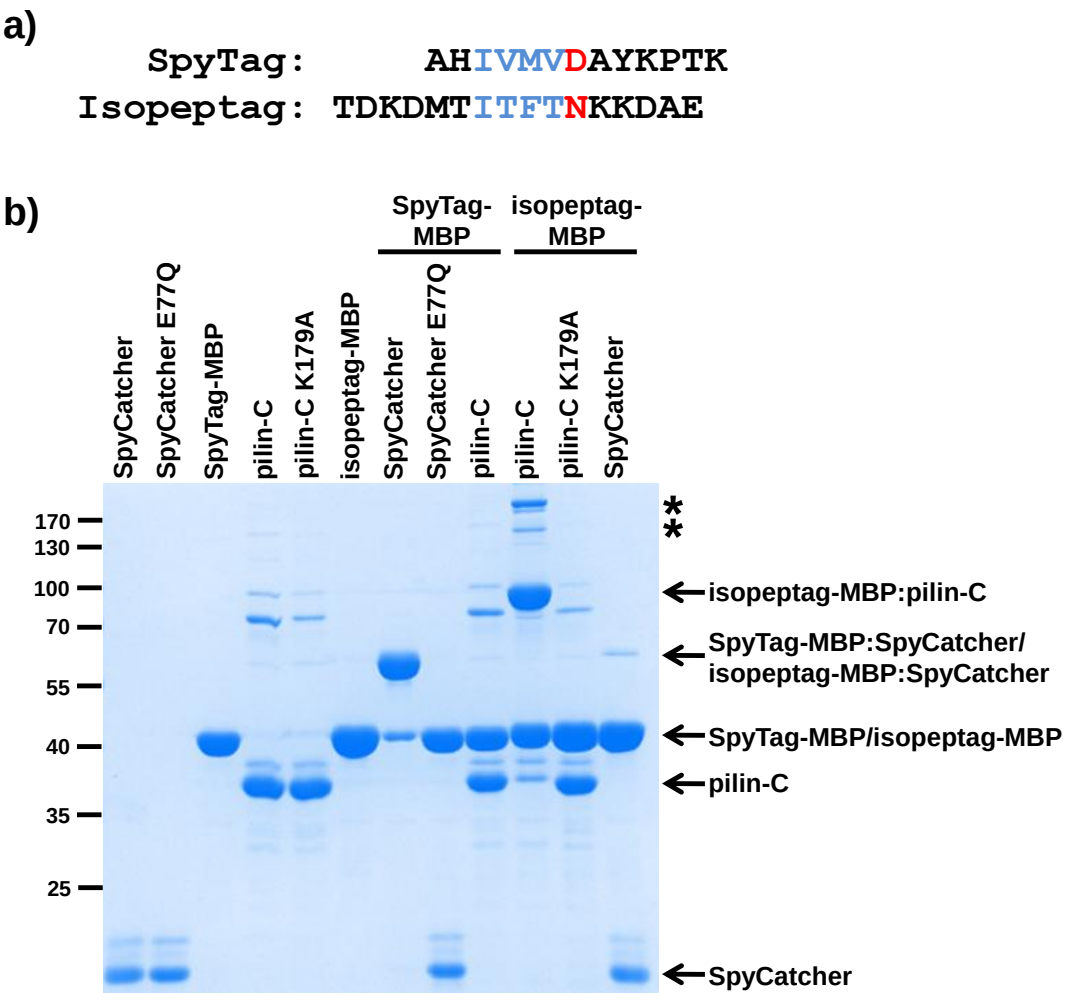


Figure 6.11. SpyTag:SpyCatcher and isopeptag:pilin-C technologies could cross-react. a) Sequence of SpyTag and isopeptag. Residues involved in isopeptide bond formation are

shown in red. Residues that have similar chemical properties and may contribute to cross-reactivity of the two technologies are shown in blue. b) A test of cross-reactivity of the SpyTag:SpyCatcher and isopeptag:pilin-C technologies. Proteins were mixed at 10 μM in 40 mM Na_2HPO_4 – 20 mM citric acid pH 7.0 and incubated for 24 hr at 25°C. Reactions were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining.
* Side-products that formed as a result of the isopeptag-MBP and pilin-C reaction.

6.12. SpyTag and SpyCatcher reconstitution on the surface of mammalian cells

SpyCatcher forms an intermolecular isopeptide bond with a reactive Asp residue in SpyTag. To demonstrate that this reaction was specific for SpyTag and not other exposed Asp residues in surrounding proteins, I sought to demonstrate that SpyTag and SpyCatcher were able to specifically reconstitute in a complex environment on the surface of mammalian cells.

To demonstrate the specific nature of the SpyTag and SpyCatcher interaction, SpyTag was displayed on the surface of HeLa cells. To do so, SpyTag was genetically linked to the N-terminal end of the Cyan Fluorescent Protein (CFP) via a 5 residue flexible linker. The construct was then inserted in to the mammalian expression vector pDisplay, which genetically linked the transmembrane domain of the Platelet-derived Growth Factor Receptor to the C-terminal end of the SpyTag-CFP construct via a myc tag linker. This produced the construct SpyTag-CFP-TM which upon expression would be inserted in to the plasma membrane with the SpyTag-CFP region being displayed in the extracellular environment (Fig. 6.12a). The membrane displayed SpyTag could then be targeted with exogenously added Cna114 I34E, which had a similar level of activity as SpyCatcher (Fig. 5.7). Reconstitution could then be detected using an antibody against the 6X His tag linked to the N-terminal end of Cna114 I34E (Fig. 6.12a).

Upon transfection in to HeLa cells, SpyTag-CFP-TM expressed well and trafficked efficiently to the plasma membrane as apparent from the CFP signal (Fig. 6.12b). To label surface displayed SpyTag, I added Cna114 I34E to the cells at 5 μM in cell growth medium

and incubated the cells for 5 min at 25°C. I then washed unbound Cna114 I34E and labelled the cells with an anti-His primary antibody and Alexa Fluor 555 labelled secondary antibody (Fig. 6.12b). Even after a 5 min incubation period, SpyTag-CFP-TM expressing cells were specifically labelled with Cna114 I34E amidst the many proteins present in the complex environment of a mammalian cell surface and in the cellular growth medium (Fig. 6.12b). As a negative control, isopeptag-CFP-TM (Fig. 3.18a) was used and although it expressed and trafficked well to the plasma membrane, it was not labelled with Cna114 I34E at 5 μM even after a 15 min incubation period (Fig. 6.12b). Therefore, the SpyTag:SpyCatcher technology could be used for specific cell surface labelling applications.

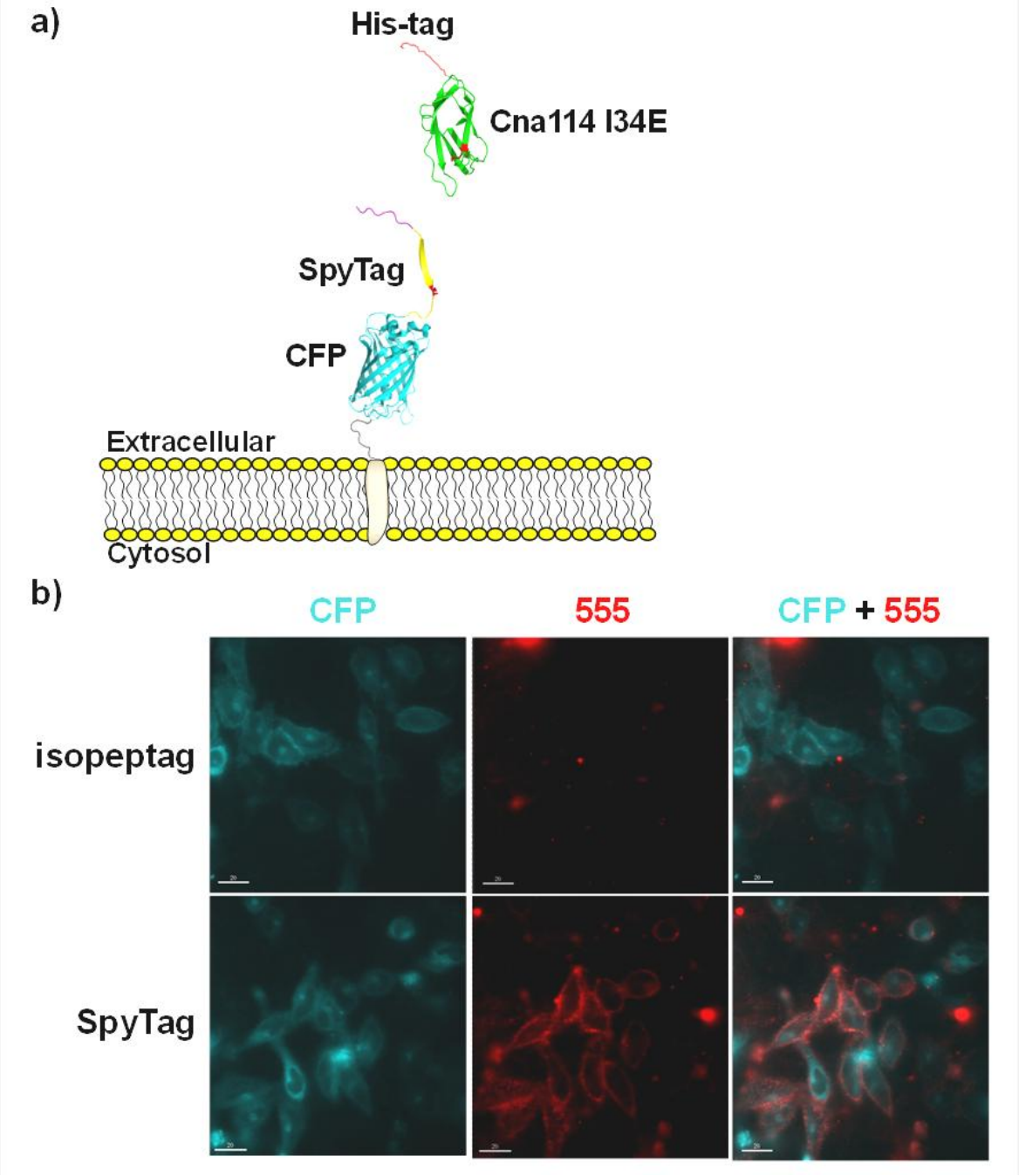


Figure 6.12. SpyTag could be specifically labelled on mammalian cells. a) A cartoon illustrating membrane anchored SpyTag-CFP-TM being targeted by exogenous Cna114 I34E. An antibody against the 6X His tag (red) at the N-terminal end of Cna114 I34E was used to detect SpyTag and Cna114 I34E reconstitution. b) Labelling of membrane anchored SpyTag with exogenous Cna114 I34E. HeLa cells expressing either isopeptag-CFP-TM or SpyTag-CFP-TM were incubated with Cna114 I34E at 5 µM in growth medium for 15 min (isopeptag-CFP-TM) or 5 min (SpyTag-CFP-TM) at 25°C. Cells were then washed and labelled with an anti-His primary antibody and an Alexa Fluor 555 labelled secondary antibody and imaged. Images are all contrast matched.

6.13. Summary

In this chapter, I investigated the characteristics of the interaction between SpyTag and SpyCatcher. Covalent reaction between the peptide-protein pair occurred rapidly and resulted in 40% depletion of SpyCatcher after one minute with a calculated rate constant of 1.4×10^3 $M^{-1}s^{-1}$. Furthermore, covalent reconstitution between the peptide-protein pair was not highly dependent on the concentration of the reactants, where similar levels of product formation were observed at 1, 5, and 10 μM of each reactant. Moreover, peptide-protein covalent reconstitution proceeded efficiently at varying protein ratios and at a range of incubation temperatures, pH values, buffer compositions, and in the presence of commonly used detergents. Covalent isopeptide bond formation between SpyTag and SpyCatcher leads to the release of one water molecule per reaction. Although water was present in large excess in the surrounding environment, the SpyTag and SpyCatcher reaction was irreversible and not susceptible to hydrolysis. Interestingly, the SpyTag:SpyCatcher and isopeptag:pilin-C pairs were able to cross-react, where isopeptag-MBP formed a covalent complex with SpyCatcher, while covalent reconstitution was not detected between SpyTag and pilin-C. This was likely a result of similarities at key residues in SpyTag and isopeptag. Additionally, I investigated cell surface labelling on mammalian cells using SpyTag and SpyCatcher to demonstrate their ability to specifically reconstitute in a complex environment. When anchored in the plasma membrane and displayed on the surface of HeLa cells, SpyTag could be specifically targeted and labelled with SpyCatcher. Therefore, the SpyTag:SpyCatcher system is a robust technology that can be used in a variety of biologically relevant experimental conditions.

Chapter 7: Summary and General Discussion

Chapter 7: Summary and General Discussion

Peptide tags are routinely fused to proteins of interest to allow the specific capture and manipulation of target proteins for a variety of biological applications including protein purification, assembly, immunological detection, and cellular imaging (Terpe, 2003; Popp et al., 2007; Appleby et al., 2009). However, there is an inherent limitation to the affinity of a peptide-protein pair as a result of the limited surface area available for binding and the utilization of weak non-covalent interactions (Houk et al., 2003). Consequently, the capture of target proteins via peptide fusion tags is reversible, which leads to reduced sensitivity of biological assays. To overcome such barriers, methods have been developed that utilize native chemical ligation (Wilken and Kent, 1998), inteins (Appleby et al., 2009), or enzymes (Abrahmsen et al., 1991; Popp et al., 2007) to covalently link two polypeptide sequences. To increase the repertoire of available peptide fusion tags, I sought to develop a new class of peptide fusion tags that could form spontaneous and irreversible covalent bonds to their corresponding protein partners. To do so, my project took inspiration from a feature present in certain outer-membrane proteins of Gram-positive bacteria that is used to stabilize protein structures, namely intramolecular isopeptide bond formation (Kang et al., 2007; Kang and Baker, 2009; Alegre-Cebollada et al., 2010; Hagan et al., 2010). I investigated dissection of intramolecular isopeptide bond containing proteins to develop a novel class of peptide-protein partners that interacted through spontaneous intermolecular isopeptide bond formation.

7.1. A general method for developing intermolecular isopeptide bond forming peptide-protein pairs

It has been well established that proteins can be genetically split in to two fragments that once expressed, can spontaneously interact in solution and refold in to an active protein

(Johnsson and Varshavsky, 1994; Ghosh et al., 2000; Zhang et al., 2004; Ozawa, 2006). Using this approach I dissected three different intramolcular isopeptide bond containing proteins to develop covalently interacting peptide-protein pairs.

7.1.1. Dissection of Spy0128

Spy0128 is the major pilin subunit of *S. pyogenes* and was the first protein identified to contain intramolecular isopeptide bonds (Kang et al., 2007). It has 2 domains (N- and C-domain) composed of all β-strands and within each domain, the first and last β-strands are linked through a Lys-Asn isopeptide bond (Kang et al., 2007). As proof-of principle studies for the generation of covalently interacting peptide-protein pairs, I split Spy0128 individually in both the N- and C-domains by removing either the first or last β-strand in the protein to generate the peptide, while the remaining protein fragment constituted the binding partner (Zakeri and Howarth, 2010). Both peptides were able to covalently reconstitute with their binding partners, although the C-domain split constructs were more efficient (Fig. 3.4 and Fig. 3.5). A possible explanation for the discrepancy between the reconstitution efficiencies of the split N- and C-domain constructs may be that isopeptag-N (28 residues) is a much larger peptide than isopeptag (16 residues), and may thus be more structured in solution by forming hair-pin structures for example. Consequently, the peptide would be in a less receptive state for recognition and reconstitution with its binding partner. Alternatively, removal of the N-terminal β-strand may be more disruptive to the structure of the resulting binding partner than removal of the C-terminal β-strand, since protein folding is likely to be initiated at the N-terminal domain first as the protein is being translated. This would result in reduced reconstitution efficiency by increasing the amount of time required for the peptide to recognize and bind to the highly dis-ordered binding partner. Optimization of the C-domain split constructs led to the development of the isopeptag:pilin-C pair (Fig. 3.9), which was able to achieve ~60% reconstitution after a 72 hr reaction (Fig. 3.11a). The reaction between the

isopeptag:pilin-C pair was robust and could proceed at a range of concentrations, temperatures, pH values, buffer compositions, and in the presence of detergent. Furthermore, isopeptag and pilin-C could be genetically encoded and specifically reconstituted inside the complex environment of a bacterial cytosol (Fig. 3.17), and they could also be used for specific cell surface labelling on the surface of mammalian cells (Fig. 3.18).

While the isopeptag:pilin-C technology validated my approach of developing a peptide-protein pair that was able to form a spontaneous isopeptide bond, reconstitution between the pair was not efficient and the protein partner was larger than desired. An attractive feature of the isopeptag:pilin-C pair was that both partners could be genetically encoded, thus allowing for *in vivo* applications of this technology. Accordingly, it would have been ideal to have a protein partner be as small as possible for fusion to a protein of interest. However, pilin-C was a large protein (31.0 kDa without the N-terminal 6X His tag), which may have limited its applicability for use as a fusion protein. Consequently, I proceeded to explore dissection of smaller isopeptide bond containing proteins to develop more efficiently interacting peptide-protein pairs.

7.1.2. Dissection of ACE19

Using a structural template of residues involved in the formation of a previously unrecognized isopeptide bond (Kang et al., 2007), I explored the Protein Data Bank for other proteins that contained previously unrecognized intramolecular isopeptide bonds. This led to the discovery of ACE19, a subdomain from a collagen binding protein in *E. faecalis* (Ponnuraj and Narayana, 2007) that contained an intramolecular isopeptide bond, which I validated using mass spectrometry (Fig. 4.4b). Subsequently, I split ACE19 and developed a peptide and two versions of the binding partner (Fig. 4.5a). Although the split ACE19 constructs were able to covalently reconstitute as determined by SDS-PAGE, the efficiency of the reaction was very low (Fig. 4.5b). A possible explanation for why the split ACE19 constructs were

much more inefficient at reconstitution than the split Spy0128 consructs may be that in ACE19 the intramolecular isopeptide bond forms between β-strands in opposite β-sheets, whereas in Spy0128 both intramolecular isopeptide bonds form between adjacent β-strands within the same β-sheet. Consequently, removal of a β-strand in ACE19 may be more disruptive to the structure of the resulting binding partner than in Spy0128, thus leading to reduced reconstitution efficiency. Alternatively, MBP was fused to the C-terminal end of ACEA which may have sterically hindered reconstitution of the peptide with its binding partner. However, this is unlikely since the C-terminal end of ACEA is exposed to solution, thus reducing the likelihood of steric clashes between MBP and ACE44 or ACE50.

7.1.3. Dissection of CnaB2

CnaB2 is a small domain of the much larger fibronectin binding protein FbaB, and it contains an intramolecular isopeptide bond that forms between the side chains of a Lys and an Asp residue (Hagan et al., 2010). Splitting of the CnaB2 scaffold led to the development of the peptide PepA and the corresponding binding partners Cna110 and Cna114, which were able to reconstitute more efficiently than the isopeptag:pilin-C pair (Fig. 4.8b). Subsequently, we used NMR spectroscopy to gain insight into the structure of the fragmented protein Cna110 in the presence and absence of the peptide binding partner. Analysis of the structure of Cna110 revealed that it was present in a partially disordered state, however in the presence of Cna peptide the peptide-protein pair were able to reconstitute and fold in to a structure resembling wild-type CnaB2 (Fig. 4.12a).

I further proceeded to try to increase the reconstitution efficiency of the split CnaB2 constructs using phage display and rational modifications. Investigation of a commercially available dodecapeptide phage display library with a diversity of 10^9 different random peptide variants did not yield any covalently binding peptides for either Cna110 or Cna114 (Table 5.1) (New England Biolabs, 2009). Consequently, I proceeded to try to enhance the reconstitution

efficiency of the split CnaB2 constructs by rationally modifying Cna114 and PepA. After conducting a series of mutations on the Cna114 scaffold, I found that mutations I34E and M69Y significantly increased reconstitution efficiency of the binding partner (Table 5.2). Therefore, I combined the mutations to make the optimized binding partner SpyCatcher. On the other hand, all attempts at modifying the sequence of PepA resulted in reduced reconstitution efficiency, while extension of the peptide sequence at its C-terminal end with residues present in the FbaB protein significantly enhanced reconstitution efficiency (Table. 5.3). Hence, the C-terminal end of PepA was extended by 5 residues to make the optimized peptide SpyTag.

The optimized SpyTag:SpyCatcher technology was far more efficient than the isopeptag:pilin-C technology. When SpyTag-MBP was mixed with SpyCatcher each at 10 μM, over 40% of SpyCatcher was depleted after 1 min and all of SpyCatcher was depleted after 2 hr (Fig. 6.3b). Reaction between SpyTag-MBP and SpyCatcher was irreversible and efficient reconstitution was observed at a range of concentrations, temperatures, pH values, buffer compositions, and in the present of detergents. Interestingly, the SpyTag:SpyCatcher and isopeptag:pilin-C technologies were able to cross-react, likely as a result of sequence similarities between SpyTag and isopeptag at key residues (Fig. 6.11b). Furthermore, the SpyTag:SpyCatcher pair were able to specifically reconstitute in the complex environment on a mammalian cell membrane (Fig. 6.12).

7.2. A new class of covalently binding peptide fusion tags

Herein, I have developed a new class of peptide fusion tags that can interact with their binding partners via irreversible and spontaneous isopeptide bond formation. The efficiently reconstituting SpyTag:SpyCatcher pair is composed of the regular 20 amino acids and can therefore be genetically encoded. Since there are no Cys residues present in either peptide-protein partner, reconstitution is not dependent on the redox status of the environment.

Furthermore, covalent reconstitution can be achieved both *in vitro* and *in vivo* under most biologically relevant conditions and without the requirement of any accessory factors. The SpyTag:SpyCatcher system represents a new class of peptide fusion tags and increases the array of molecular tools that are available to scientists, where its unique attributes make it advantageous for use under certain experimental conditions.

However, there are several limitations to the SpyTag:SpyCatcher system. Reconstitution is not traceless and when SpyTag and SpyCatcher are used as fusion constructs to link two polypeptide sequences, a 13.5 kDa protein will remain in between. The rate constant for the second order reaction between SpyTag-MBP and SpyCatcher was 1.4 x 10^3 $M^{-1}s^{-1}$ (Fig. 6.3c), which is far below the diffusion limited reaction rates achieved by some highly efficient enzymes with second order rate constants of 10^8-10^{10} $M^{-1}s^{-1}$ (Stroppolo et al., 2001). Therefore, there is significant room for improvement of the reconstitution efficiency of the SpyTag:SpyCatcher pair. Furthermore, the structurally similar Cna110 protein was present in a disordered state, and therefore when SpyCatcher is fused to a protein of interest it may cause solubility and folding issues.

7.3. A new class of split proteins

It has been well established for over half a century that certain split proteins are able to interact spontaneously and refold to regain the original protein function (Richards, 1958). This elegant technique has been successfully applied to many protein scaffolds including enzymes and fluorescent proteins to develop sensors for the study of protein-protein interactions (Ozawa, 2006). In these studies, the objectives were to split reporter proteins in such a way that the two fragments could not reconstitute spontaneously, but would only reconstitute when brought in to close proximity (Fig. 7.1a). Subsequently, the split reporter protein fragments would be fused to two proteins of interest and if these proteins interacted, then the split reporter protein fragments would be brought in to close proximity, resulting in

reconstitution and the generation of a measurable signal (Fig. 7.1a) (Ozawa, 2006; Ghosh et al., 2000; Johnsson and Varshavsky, 1994; Singh et al., 2006).

In this study, protein splitting was used to develop a new class of biotechnological tool. Three different intramolecular isopeptide bond containing proteins were split by removing a terminal β-strand to demonstrate a general approach for developing covalently interacting peptide-protein pairs. Unlike split reporter proteins that were designed to reconstitute only if their fusion protein partners interacted, the SpyTag:SpyCatcher system was designed to covalently link two fusion proteins that would otherwise not interact (Fig. 7.1b). This represented a new class of split proteins that were able to spontaneously reconstitute and form an irreversible isopeptide bond. Furthermore, splitting of Spy0128, ACE19, and CnaB2 revealed that the choice of scaffold was very important to the reconstitution efficiency of the split fragments. Split CnaB2 constructs (Fig. 6.1b) were the most efficient at reconstitution followed by split Spy0128 constructs (Fig. 3.4c and Fig. 3.9b), and then split ACE19 constructs (Fig. 4.5b), which showed only minor amounts of covalent product formation. However, this was based on a strategy of splitting proteins by removing a terminal β-strand and reconstitution efficiencies of the scaffolds may have been different if the proteins were split at different positions.

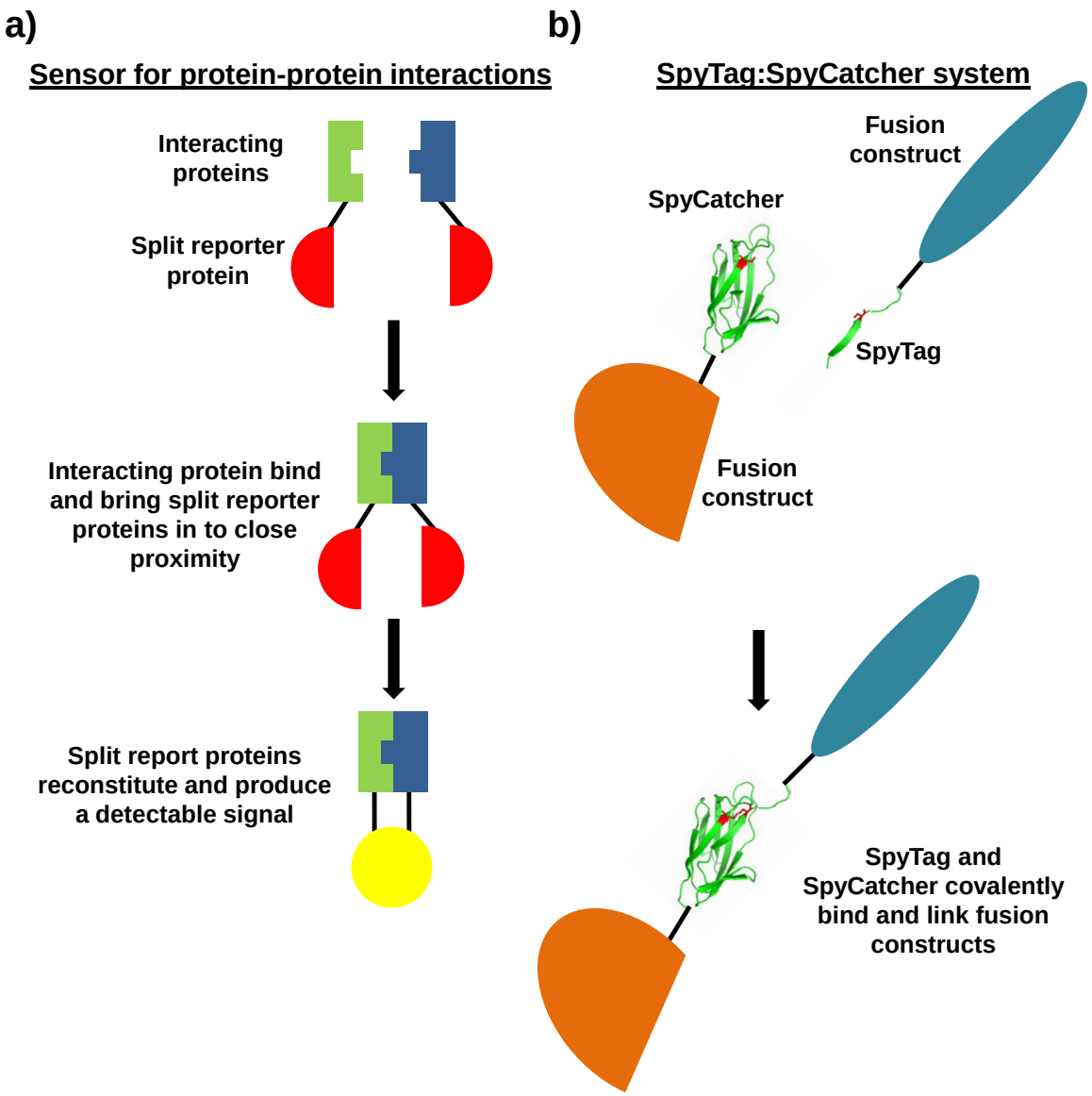


Figure 7.1. Split protein technologies. a) Reporter proteins can be split to develop sensors for protein-protein interactions. When they are present in solution the split reporter proteins cannot reconstitute. However, if their fusion proteins interact then the two fragments of the reporter protein would be positioned in close proximity which would induce reconstitution to form an active reporter protein that would then produce a detectable signal to signify a protein-protein interaction event. b) SpyTag and SpyCatcher can each be linked to a fusion protein, and when they are present in solution they can spontaneously reconstitute via an isopeptide bond, resulting in a covalent linkage of the two fusion proteins. Isopeptide bond forming residues are shown in red.

7.4. Structural insights in to split protein reconstitution

We performed preliminary NMR studies of covalent reconstitution between Cna110 and Cna peptide. The split Cna110 protein was present in a disordered state, where the hydrophobic core of the protein was disturbed (Fig. 4.12a). Yet in the presence of Cna peptide the hydrophobic core of Cna110 was restored and Cna110 was able to fold to a structure similar to wild-type CnaB2 (Fig. 4.12a).

It would be valuable to conduct detailed NMR structural studies on the SpyTag:SpyCatcher pair to gain a deeper understanding of the structural aspects of split protein reconstitution. SpyCatcher reconstitution efficiency was enhanced by the introduction of I34E and M69Y mutations in Cna114 based on rational hypotheses. However, we cannot be certain of the effects that these modifications had on the Cna114 structure until we have detailed structural information on Cna114 and SpyCatcher. The I34E mutation was made to increase the solubility of a flexible loop near the SpyTag binding site, based on the hypothesis that the flexible loop may have folded in to the SpyTag binding pocket and interacted with the exposed hydrophobic core of the split protein (Fig. 5.3b). Yet Cna110 (and likely SpyCatcher) lacked a defined structure in solution. Thus, it is difficult to imagine how the I34E mutation would aid reconstitution between a disordered protein and a likely unstructured peptide. Furthermore, the M69Y mutation was made to disrupt potential transient hydrophobic interactions between Cna114 molecules that may have interfered with peptide-protein interactions (Fig. 5.3c). This hypothesis was based on the assumption that the small CnaB2 domain in the large FbaB protein may interact with other nearby protein domains via hydrophobic interactions. Although the M69Y mutation was successful in enhancing the reconstitution efficiency of Cna114, it has yet to be determined how the M69Y mutation affected the structure of Cna114 in solution. It may be possible to determine whether Cna114 forms dimers or larger protein aggregates in solution using dynamic light scattering (DLS) or gel filtration techniques. In the past, I had attempted to determine

potential interactions of Cna114 molecules in solution using DLS, however due to problems with the instrumentation I could not obtain reproducible results. Furthermore, a structure of the SpyTag:SpyCatcher complex would provide valuable information regarding how the C-terminal extension of PepA with YKPTK residues led to a dramatic increase in reconstitution efficiency. It may be possible that the two positively charged Lys residues in SpyTag may be interacting with a cluster of negatively charged residues nearby on SpyCatcher (Fig. 5.4). However, this conclusion needs to be validated with structural information. A better understanding of how the modifications made to Cna114 and PepA led to such a dramatic enhancement in reconstitution efficiency could be applied to further enhance the speed of reconstitution between SpyTag and SpyCatcher.

7.5. Insights in to intramolecular isopeptide bond formation

Intramolecular isopeptide bonds are thought to form spontaneously as proteins naturally fold (Hu et al., 2010). As such, the proposed mechanism for isopeptide bond formation has been based on data from proteins in which isopeptide bond formation has reached completion (Kang et al., 2007; Helgstrand et al., 2003; Hagan et al., 2010; Hu et al., 2010; Kang and Baker, 2011). The SpyTag:SpyCatcher system can be utilized to better understand the mechanism of autocatalytic isopeptide bond formation.

The covalent peptide-protein pairs developed in this study have provided valuable insights in to the autocatalytic isopeptide bond formation reaction. Spontaneous intermolecular isopeptide bond formation between the isopeptag:pilin-C and SpyTag:SpyCatcher pairs demonstrated that the isopeptide bond forming reaction was robust and could occur efficiently from 4-37°C and at pH 5-8. Furthermore, the rational modifications investigated to optimize the reconstitution efficiency of the split CnaB2 constructs revealed that the residues in the C-terminal β-strand were critical for efficient isopeptide bond formation. More importantly, the efficiency of intermolecular isopeptide bond formation was

enhanced by modifying PepA and Cna114 at locations that were likely not directly involved in autocatalytic isopeptide bond formation. Therefore, it was likely that in the bimolecular reaction the limiting factor was binding and refolding of the peptide-protein pair, while the isopeptide bond forming reaction proceeded rapidly. To better understand the mechanism of isopeptide bond formation, structural studies could be performed on SpyTag and SpyCatcher in solution using NMR spectroscopy. Since SpyCatcher is likely in a disordered state in solution, the addition of SpyTag would allow for monitoring of split protein refolding, and also spontaneous isopeptide bond formation during the protein folding event.

7.6. Further enhancing the reconstitution efficiency of SpyTag and SpyCatcher

We believe that it is possible to further enhance the speed of reconstitution between SpyTag and SpyCatcher, to possibly achieve reaction rates that approach the diffusion limit. Although it appears that the sequence of the C-terminal β-strand cannot be altered in SpyTag, it may be possible to further enhance the reconstitution efficiency of the peptide-protein pair by introducing favourable interactions outside of the peptide binding pocket, similar to the C-terminal extension of PepA which led to the development of SpyTag. To conduct such experiments it would be advantageous to explore diverse peptide and protein libraries, which can be achieved using phage display. The phage display experiments performed in this study were based on a random peptide library, which did not preserve the core residues in PepA required for efficient isopeptide bond formation. In future phage display experiments, only peripheral residues not directly involved in isopeptide bond formation in both SpyTag and SpyCatcher would be randomized to try to enhance the affinity of the peptide-protein pair. By this approach, it may be possible to isolate a covalently binding peptide-protein pair whose reaction is diffusion limited.

7.7. Potential applications of the SpyTag:SpyCatcher technology

7.7.1. Signal amplification of binding events

An application for which the SpyTag:SpyCatcher system may be advantageous would be for the development of highly sensitive assays via signal amplification for binding events. To develop highly sensitive protein nanowires, Men at al. made fusion constructs of an amyloid protein linked to either a binding protein or an enzyme that could produce a measurable signal (Men et al., 2009). When mixed at an optimized ratio, the amyloid proteins self-assembled to produce fibrils that contained several signal producing enzymes for every binding protein in the fibril, which they termed nanowires (Men et al., 2009). When tested in an ELISA assay, the nanowires were 100-fold more sensitive than the convention detection method of a primary antibody conjugated to a signal producing enzyme, which had a binding to signal enzyme ratio of 1:1 (Men et al., 2009). The increased sensitivity of the nanowires was a result of the presence of a large number of signal producing enzymes for every binding event (Men et al., 2009).

Using a similar approach, SpyTag and SpyCatcher could be used to increase the sensitivity of biological assays. If SpyTag were to be linked at the N-terminal end of SpyCatcher, the resulting fusion construct would be able to spontaneously self-assemble to produce a linear chain extending many molecules (Fig. 7.2a). This system could be used to increase the sensitivity of a conventional ELISA experiment. If SpyTag were to be fused to a binding molecule such as an antibody and the antibody-SpyTag complex was allowed to first bind the antigen, then dye labelled SpyTag-SpyCatcher fusions could be added to produce an enhanced signal for the binding event, thereby increasing the sensitivity of the assay (Fig. 7.2b). However, a limitation of this approach would be that the SpyTag-SpyCatcher fusion construct could in theory reconstitute uncontrollably, creating extremely long protein chains. Control of the length of the self-assembled chain could possibly be achieved by adding free SpyTag or SpyCatcher at a concentration that is optimized to achieve a certain chain length.

Furthermore, it would be difficult to purify un-polymerized SpyTag-SpyCatcher subunits. It may be possible to express and purify an unreactive version of SpyTag-SpyCatcher, by adding a fusion protein that inhibits the reactivity of SpyTag in the SpyTag-SpyCatcher subunit. The inhibitive fusion protein could then be cleaved using proteases to activate the SpyTag-SpyCatcher subunits.

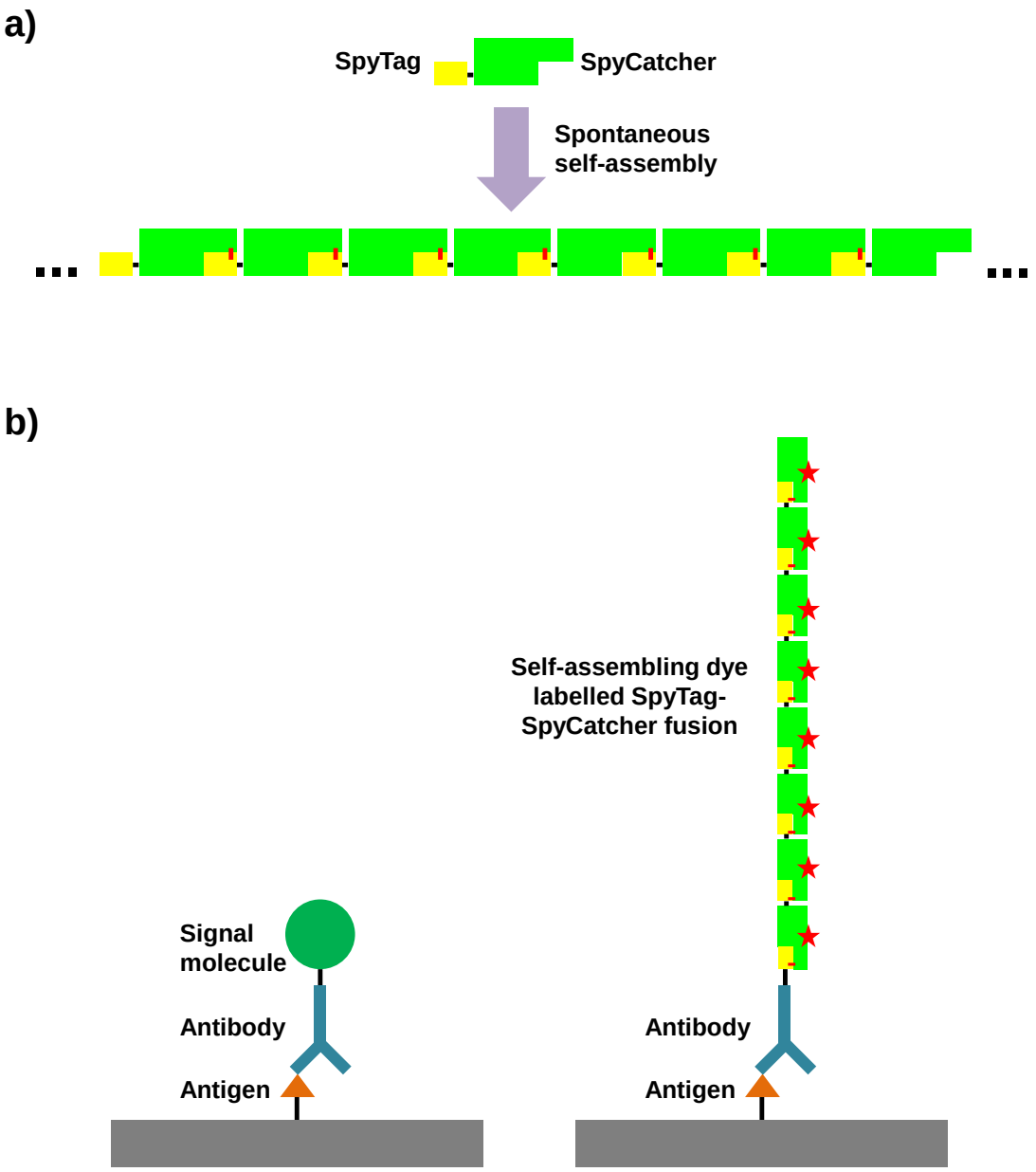


Figure 7.2. Self-assembling of SpyTag-SpyCatcher fusions. a) Cartoon illustrating a self-assembling SpyTag-SpyCatcher fusion. SpyTag can be linked to the N-terminal end of SpyCatcher with a linker in between, and the resulting fusion construct would be able to spontaneously assemble to form a covalently linked linear chain of SpyTag-SpyCatcher molecules. Red bars represent an isopeptide bond. b) A cartoon illustrating an ELISA experiment where SpyTag-SpyCatcher fusions can be used to achieve enhanced sensitivity (similar to nanowires produced by (Men et al., 2009)). Left: A conventional ELISA experiment, where an antigen linked to a solid support can be bound by an antibody conjugated to a signal producing molecule. For every binding molecule there is one signal molecule. Right: An ELISA experiment, where an antigen linked to a solid support can be bound by an antibody linked to SpyTag. After antibody binding, dye labelled (red star) SpyTag-SpyCatcher fusions can be added to bind the antibody-SpyTag construct and produce an enhanced signal, where for every binding event there are many signalling molecules present. The

length of the self-assembled chain can be controlled by adding SpyCatcher not linked to SpyTag.

7.7.2. Assembly of novel protein architectures

An intriguing potential application for the SpyTag:SpyCatcher technology would be the assembly of novel protein architectures. A commonly used method of assembling proteins is through the use of coiled coils (Ghosh et al., 2000), however the SpyTag:SpyCatcher technology would bring a new dimension of irreversibility. Since SpyTag and SpyCatcher can be genetically fused to proteins of interest, the peptide-protein pair can act as building blocks to covalently link fusion proteins for the development of multi-protein complexes (Fig. 7.3). SpyTag could be possibly positioned in unstructured regions of proteins such as flexible loops or the linker sequences between proteins, and subsequently targeted by SpyCatcher fused to a protein of interest or a detectable chemical moiety (Fig. 7.3). The irreversible covalent bond formed between the peptide-protein pair would allow for stable localization of proteins and the formation of branched protein complexes. Similarly, the SpyTag:SpyCatcher pair could be used for the development of protein locks, protein polymers, or for the co-localization of proteins or enzymes (Fig. 7.3). As such, SpyTag and SpyCatcher could simply provide a means of spontaneously linking proteins via an isopeptide bond without the requirement of any accessory factors.

However, before the SpyTag:SpyCatcher technology can be used for such applications as outlined in Fig. 7.3, several concerns must be addressed. Insertion of the SpyTag sequence in linker regions between proteins or in flexible loops may result in reduced reconstitution efficiency with SpyCatcher due to steric hindrances, when compared to localization of SpyTag at the terminal ends of a protein. Also, the display of SpyTag in solution may be reduced when the peptide is not terminally located, thereby leading to reduced reconstitution efficiency with SpyCatcher. It may be possible to use long linker sequences to increase the display of SpyTag located in unstructured protein regions.

Furthermore, the linkage of several SpyCatcher molecules within one polypeptide chain may result in differences in reconstitution efficiency of the individual SpyCatcher molecules, depending on the position of each monomer in the chain. Finally, the use of SpyTag and SpyCatcher to assemble large protein complexes may result in undesired side-reactions occurring, where proteins of interest would be linked in random configurations. To overcome these limitations, extensive optimization of the binding conditions would be required for each experiment.

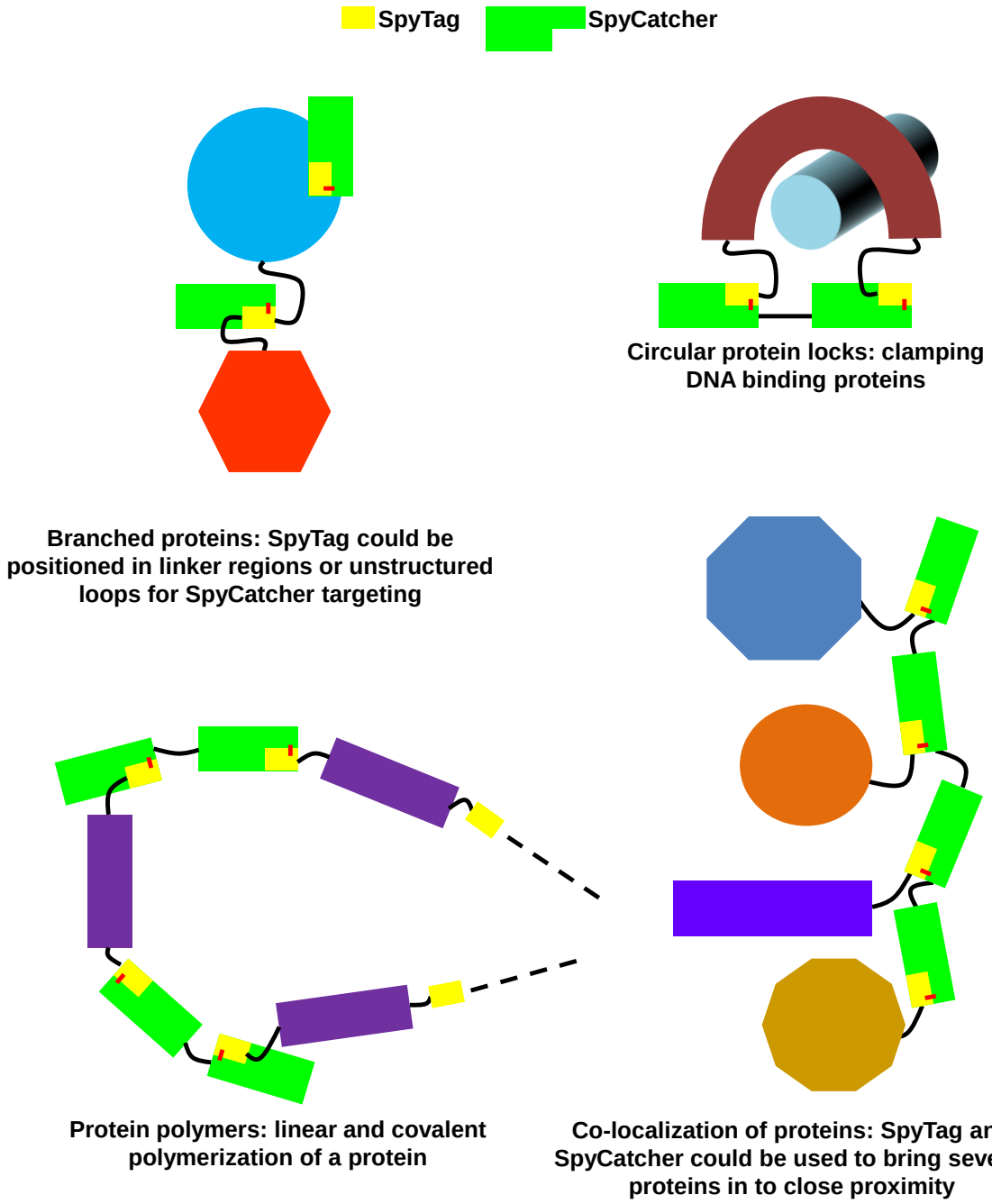


Figure 7.3. Potential applications of SpyTag and SpyCatcher for the assembly of novel protein architectures. Cartoons illustrating various ways that SpyTag and SpyCatcher could be linked to proteins of interest to assemble large multi-protein complexes.

7.7.3. Development of a series of isopeptide bond forming peptide fusion tags

In this study, two covalent peptide tagging systems were developed, isopeptag:pilin-C and SpyTag:SpyCatcher. However, it would be advantageous to dissect other intramolecular isopeptide bond containing proteins to develop similar tagging system, where each peptide tag would be highly specific for its particular binding partner. In doing so, there would be a series of isopeptide tagging systems available that could be used to engineer novel protein architectures similar to those depicted in Fig. 7.3, however with the added benefit of greater control over structural assembly of protein complexes. By using several tagging systems simultaneously, one would be able to direct the location and the orientation of assembly of different target proteins.

Herein, I split three intramolecular isopeptide bond containing proteins to develop four covalent peptide-protein pairs. Three of the isopeptide bonds dissected formed between Lys and Asn side chains (two isopeptide bonds in Spy0128 and one in ACE19), while the fourth isopeptide bond dissected formed between Lys and Asp side chains (CnaB2). Interestingly, the peptide-protein pair that reconstituted via a Lys-Asp isopeptide bond was far more efficient than the Lys-Asn reconstituting pairs. It may have been that intermolecular isopeptide bond formation between Lys-Asp residues was more efficient than Lys-Asn residues. However, this was unlikely to be the reason for the observed discrepancy in the reconstitution efficiencies of the peptide-protein pairs. As discussed in section 7.5, the limiting factor in intermolecular isopeptide bond formation was likely recognition and folding of the peptide-protein pair. Therefore, it was likely that the split CnaB2 peptide-protein pair had a higher affinity for each other than the split Spy0128 and ACE19 peptide-protein pairs.

From only three split protein scaffolds it is difficult to draw specific conclusions regarding the type of protein that would be ideal to split for the development of an efficient covalently reconstituting peptide-protein pair. It is likely that dissection of CnaB-type domains where an isopeptide bond links two adjacent β-strands (first and last β-strands in the domain)

would result in more efficiently reconstituting peptide-protein pairs (Kang and Baker, 2011), as observed with Spy0128 and CnaB2. Splitting of CnaA-type domains, where two β-sheets are linked by an isopeptide bond that forms between the first and penultimate β-strands (Kang and Baker, 2011) as in the ACE19 protein, would likely result in a greater level of structural perturbation and hence reduced reconstitution efficiency. However, more intramolecular isopeptide bond containing proteins must be split from both CnaA and CnaB-type domains to better understand what type of scaffolds would produce efficiently reconstituting peptide-protein pairs. Furthermore, for certain applications such as Western blot (discussed in section 7.7.4) and cellular imaging experiments, the size of the binding partner is not an issue. Therefore, if larger split isopeptide bond containing proteins can reconstitute faster than SpyTag and SpyCatcher, then they would be very valuable for such applications.

7.7.4. The SpyTag:SpyCatcher system might be used as a tool for sensitive detection of target proteins

Many commonly used peptide fusion tags such as the 6X His tag, c-myc-tag, and HA tag can be detected with specific antibodies for various biological applications including Western blot, ELISA, and cellular imaging (Terpe, 2003). In this study, I have demonstrated the applicability of the SpyTag:SpyCatcher technology for cellular imaging. However, it still remains to be demonstrated whether this technology can be used as a more efficient, stable, and economical alternative to antibodies for the detection of target proteins in certain other biological experiments. For applications such as Western blot and ELISA, SpyCatcher could be labelled with a chemical probe or conjugated to a signal molecule such as horseradish peroxidase (HRP) and used to detect SpyTag fusion proteins. The irreversible SpyTag:SpyCatcher covalent bond would be advantageous over non-covalent antibody-antigen interactions, where it would allow for more stringent wash steps to minimize

background signals for experiments where this is a concern. Also, the fast SpyTag:SpyCatcher reaction would be amenable for time consuming experiments such as Western blots. Additionally, the SpyTag:SpyCatcher system would provide a cost advantage over the use of antibodies, which are very expensive. SpyCatcher (for subsequent dye labelling) or SpyCatcher conjugates could be expressed in *E. coli* and purified using conventional techniques at a minimal cost per unit compared to commercially purchased antibodies against commonly used peptide fusion tags. However, further studies are required to evaluate the utility of the SpyTag:SpyCatcher technology for use in Western blot and ELISA experiments.

7.7.5. The SpyTag:SpyCatcher system might be used to tether target proteins to a solid support

To screen a protein of interest against a library of compounds, phage, or any potential ligands, it is often tethered to a solid support such as magnetic beads. To covalently couple proteins to magnetic beads, specific cross-linkers are used that generally react with a nucleophilic functional group on the magnetic bead and a terminal nucleophilic residue on the protein. However, as an alternative if commercial preparations of magnetic beads can be purchased pre-assembled with SpyCatcher (which can be made at low cost), then any SpyTag fusion protein can be quickly coupled to the magnetic beads under physiological conditions by simply mixing the proteins with the magnetic beads without the requirement of any accessory factors. This would result in an irreversible covalent linkage of the proteins to the magnetic beads. For elution of proteins from the beads a protease cleavage site can be positioned between SpyTag and the protein to which it is fused.

7.8. Concluding remarks

Herein, I developed a new and efficient covalent tagging system in SpyTag and SpyCatcher, which has potential for use in a variety of biological applications. The SpyTag:SpyCatcher system is a versatile molecular tool that utilizes spontaneous isopeptide bond formation to irreversibly link two polypeptide sequences under most biologically relevant experimental conditions. Findings from this study have contributed to: (i) the development of a new class of peptide fusion tags, (ii) the field of split protein technology, (iii) a better understanding of the experimental parameters under which autocatalyzed isopeptide bond formation can proceed, (iv) an understanding of the approach required to enhance covalent peptide-protein reconstitution efficiency, (v) structural insights in to split protein reconstitution, and (vi) the development of a tool to allow mechanistic studies of spontaneous isopeptide bond formation. Future studies will be focused on applications of the SpyTag:SpyCatcher technology for diverse biological experiments, further enhancement of the reconstitution efficiency of the SpyTag:SpyCatcher pair using phage display, and the development of a family of isopeptide-dependent peptide-protein pairs.

References

Abrahmsen,L., Tom,J., Burnier,J., Butcher,K.A., Kossiakoff,A., and Wells,J.A. (1991). Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. Biochemistry *30*, 4151-4159.

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

Appleby,J.H., Zhou,K., Volkmann,G., and Liu,X.Q. (2009). Novel Split Intein for trans-Splicing Synthetic Peptide onto C Terminus of Protein. J. Biol. Chem. *284*, 6194-6199.

Atwell,S. and Wells,J.A. (1999). Selection for improved subtiligases by phage display. Proc. Natl. Acad. Sci. U. S. A *96*, 9497-9502.

Bakker,E.N., Pistea,A., and VanBavel,E. (2008). Transglutaminases in vascular biology: relevance for vascular remodeling and atherosclerosis. J. Vasc. Res. *45*, 271-278.

Beckett,D., Kovaleva,E., and Schatz,P.J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci. *8*, 921-929.

Beligere,G.S. and Dawson,P.E. (1999). Conformationally Assisted Protein Ligation Using C-Terminal Thioester Peptides. J. Am. Chem. Soc. *121*, 6332-6333.

Braisted,A.C., Judice,J.K., and Wells,J.A. (1997). Synthesis of proteins by subtiligase. Methods Enzymol. *289*, 298-313.

Budzik,J.M., Poor,C.B., Faull,K.F., Whitelegge,J.P., He,C., and Schneewind,O. (2009). Intramolecular amide bonds stabilize pili on the surface of bacilli. Proc. Natl. Acad. Sci. U. S. A *106*, 19992-19997.

Chang,T.K., Jackson,D.Y., Burnier,J.P., and Wells,J.A. (1994). Subtiligase: a tool for semisynthesis of proteins. Proc. Natl. Acad. Sci. U. S. A *91*, 12544-12548.

Chen,I., Howarth,M., Lin,W., and Ting,A.Y. (2005). Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. Nat. Methods *2*, 99-104.

Craig,L., Pique,M.E., and Tainer,J.A. (2004). Type IV pilus structure and bacterial pathogenicity. Nat. Rev. Microbiol. *2*, 363-378.

Cull,M.G. and Schatz,P.J. (2000). Biotinylation of proteins in vivo and in vitro using small peptide tags. Methods Enzymol. *326*, 430-440.

Dawson,P.E. and Kent,S.B. (2000). Synthesis of native proteins by chemical ligation. Annu. Rev. Biochem. *69*, 923-960.

de Prat,G.G. and Fersht,A.R. (1994). Generation of a family of protein fragments for structure-folding studies. 1. Folding complementation of two fragments of chymotrypsin inhibitor-2 formed by cleavage at its unique methionine residue. Biochemistry *33*, 7957-7963.

Deivanayagam,C.C., Rich,R.L., Carson,M., Owens,R.T., Danthuluri,S., Bice,T., Hook,M., and Narayana,S.V. (2000). Novel fold and assembly of the repetitive B region of the Staphylococcus aureus collagen-binding surface protein. Structure. *8*, 67-78.

Duda,R.L. (1998). Protein chainmail: catenated protein in viral capsids. Cell *94*, 55-60.

Duda,R.L., Hempel,J., Michel,H., Shabanowitz,J., Hunt,D., and Hendrix,R.W. (1995). Structural transitions during bacteriophage HK97 head assembly. J. Mol. Biol. *247*, 618-635.

Dutton,R.J., Boyd,D., Berkmen,M., and Beckwith,J. (2008). Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. Proc. Natl. Acad. Sci. U. S. A *105*, 11933-11938.

El,M.L., Terrasse,R., Dessen,A., Vernet,T., and Di Guilmi,A.M. (2010). Stability and assembly of pilus subunits of Streptococcus pneumoniae. J. Biol. Chem. *285*, 12405-12415.

Emsley,P., Lohkamp,B., Scott,W.G., and Cowtan,K. (2010). Features and development of Coot. Acta Crystallogr. D. Biol. Crystallogr. *66*, 486-501.

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.

Ghosh,I., Hamilton,A.D., and Regan,L. (2000). Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. J. Am. Chem. Soc. *122*, 5658-5659.

Green,N.M. (1990). Avidin and streptavidin. Methods Enzymol. *184*, 51-67.

Haering,C.H., Farcas,A.M., Arumugam,P., Metson,J., and Nasmyth,K. (2008). The cohesin ring concatenates sister DNA molecules. Nature *454*, 297-301.

Hagan,R.M., Bjornsson,R., McMahon,S.A., Schomburg,B., Braithwaite,V., Buhl,M., Naismith,J.H., and Schwarz-Linek,U. (2010). NMR spectroscopic and theoretical analysis of a spontaneously formed Lys-Asp isopeptide bond. Angew. Chem. Int. Ed Engl. *49*, 8421-8425.

Helgstrand,C., Wikoff,W.R., Duda,R.L., Hendrix,R.W., Johnson,J.E., and Liljas,L. (2003). The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution. J. Mol. Biol. *334*, 885-899.

Hendrickx,A.P., Budzik,J.M., Oh,S.Y., and Schneewind,O. (2011). Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds. Nat. Rev. Microbiol. *9*, 166-176.

Houk,K.N., Leach,A.G., Kim,S.P., and Zhang,X. (2003). Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. Angew. Chem. Int. Ed Engl. *42*, 4872-4897.

Howarth,M. and Ting,A.Y. (2008). Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. Nat. Protoc. *3*, 534-545.

Hu,X., Hu,H., Melvin,J.A., Clancy,K.W., McCafferty,D.G., and Yang,W. (2010). Autocatalytic Intramolecular Isopeptide Bond Formation in Gram-Positive Bacterial Pili: A QM/MM Simulation. J. Am. Chem. Soc.

Izore,T., Contreras-Martel,C., El,M.L., 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.

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.

Kadokura,H. and Beckwith,J. (2010). Mechanisms of oxidative protein folding in the bacterial cell envelope. Antioxid. Redox. Signal. *13*, 1231-1246.

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.

Kang,H.J., Paterson,N.G., Gaspar,A.H., Ton-That,H., and Baker,E.N. (2009). The Corynebacterium diphtheriae shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. Proc. Natl. Acad. Sci. U. S. A *106*, 16967-16971.

Kay,B.K., Kasanov,J., and Yamabhai,M. (2001). Screening phage-displayed combinatorial peptide libraries. Methods *24*, 240-246.

Kehoe,J.W. and Kay,B.K. (2005). Filamentous phage display in the new millennium. Chem. Rev *105*, 4056-4072.

Kleywegt,G.J. (1999). Recognition of spatial motifs in protein structures. J. Mol. Biol. *285*, 1887-1897.

Kline,K.A., Dodson,K.W., Caparon,M.G., and Hultgren,S.J. (2010). A tale of two pili: assembly and function of pili in bacteria. Trends Microbiol. *18*, 224-232.

Krauss,N., Wessner,H., Welfle,K., Welfle,H., Scholz,C., Seifert,M., Zubow,K., Ay,J., Hahn,M., Scheerer,P., Skerra,A., and Hohne,W. (2008). The structure of the anti-c-myc antibody 9E10 Fab fragment/epitope peptide complex reveals a novel binding mode dominated by the heavy chain hypervariable loops. Proteins *73*, 552-565.

Krishnan,V., Gaspar,A.H., Ye,N., Mandlik,A., Ton-That,H., and Narayana,S.V. (2007). An IgG-like domain in the minor pilin GBS52 of Streptococcus agalactiae mediates lung epithelial cell adhesion. Structure. *15*, 893-903.

Lancefield,R.C. and Dole,V.P. (1946). The properties of T antigens extracted from Group A hemolytic Streptococci. J. Exp. Med. *84*, 449-471.

Larson,M.R., Rajashankar,K.R., Crowley,P.J., Kelly,C., Mitchell,T.J., Brady,L.J., and Deivanayagam,C. (2011). Crystal Structure of the C-terminal Region of Streptococcus mutans Antigen I/II and Characterization of Salivary Agglutinin Adherence Domains. J. Biol. Chem. *286*, 21657-21666.

Liu,Q., Ponnuraj,K., Xu,Y., Ganesh,V.K., Sillanpaa,J., Murray,B.E., Narayana,S.V., and Hook,M. (2007). The Enterococcus faecalis MSCRAMM ACE binds its ligand by the Collagen Hug model. J. Biol. Chem. *282*, 19629-19637.

Lockless,S.W. and Muir,T.W. (2009). Traceless protein splicing utilizing evolved split inteins. Proc. Natl. Acad. Sci. U. S. A *106*, 10999-11004.

Mandlik,A., Swierczynski,A., Das,A., and Ton-That,H. (2008). Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends Microbiol. *16*, 33-40.

Men,D., Guo,Y.C., Zhang,Z.P., Wei,H.p., Zhou,Y.F., Cui,Z.Q., Liang,X.S., Li,K., Leng,Y., You,X.Y., and Zhang,X.E. (2009). Seeding-Induced Self-assembling Protein Nanowires Dramatically Increase the Sensitivity of Immunoassays. Nano Letters *9*, 2246-2250.

Meyer,Y., Buchanan,B.B., Vignols,F., and Reichheld,J.P. (2009). Thioredoxins and glutaredoxins: unifying elements in redox biology. Annu. Rev. Genet. *43*, 335-367.

Mishra,A., Devarajan,B., Reardon,M.E., Dwivedi,P., Krishnan,V., Cisar,J.O., Das,A., Narayana,S.V., and Ton-That,H. (2011). Two Autonomous Structural Modules in the Fimbrial Shaft Adhesin FimA Mediate Actinomyces Interactions with Streptococci and Host Cells during Oral Biofilm Development. Mol. Microbiol.

Molinari,G. and Chhatwal,G.S. (1999). Streptococcal invasion. Curr. Opin. Microbiol. *2*, 56-61.

Mora,M., Bensi,G., Capo,S., Falugi,F., Zingaretti,C., Manetti,A.G., Maggi,T., Taddei,A.R., Grandi,G., and Telford,J.L. (2005). Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc. Natl. Acad. Sci. U. S. A *102*, 15641-15646.

Nettleship,J.E., Brown,J., Groves,M.R., and Geerlof,A. (2008). Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. Methods Mol. Biol. *426*, 299-318.

New England Biolabs. Ph.D.™ Phage Display Libraries. 1.0, 1-41. 2009.
Ref Type: Pamphlet

Nylander,A., Forsgren,N., and Persson,K. (2011). Structure of the C-terminal domain of the surface antigen SpaP from the caries pathogen Streptococcus mutans. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. *67*, 23-26.

Oke,M., Carter,L.G., Johnson,K.A., Liu,H., McMahon,S.A., Yan,X., Kerou,M., Weikart,N.D., Kadi,N., Sheikh,M.A., Schmelz,S., Dorward,M., Zawadzki,M., Cozens,C., Falconer,H.,

Powers,H., Overton,I.M., van Niekerk,C.A., Peng,X., Patel,P., Garrett,R.A., Prangishvili,D., Botting,C.H., Coote,P.J., Dryden,D.T., Barton,G.J., Schwarz-Linek,U., Challis,G.L., Taylor,G.L., White,M.F., and Naismith,J.H. (2010). The Scottish Structural Proteomics Facility: targets, methods and outputs. J. Struct. Funct. Genomics *11*, 167-180.

Otaki,J.M., Tsutsumi,M., Gotoh,T., and Yamamoto,H. (2010). Secondary structure characterization based on amino acid composition and availability in proteins. J. Chem. Inf. Model. *50*, 690-700.

Ozawa,K., Wu,P.S., Dixon,N.E., and Otting,G. (2006). 15N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. FEBS J. *273*, 4154-4159.

Ozawa,T. (2006). Designing split reporter proteins for analytical tools. Anal. Chim. Acta *556*, 58-68.

Pickart,C.M. and Fushman,D. (2004). Polyubiquitin chains: polymeric protein signals. Curr. Opin. Chem. Biol. *8*, 610-616.

Pointon,J.A., Smith,W.D., Saalbach,G., Crow,A., Kehoe,M.A., and Banfield,M.J. (2010). A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. J. Biol. Chem.

Ponnuraj,K. and Narayana,S.V. (2007). Crystal structure of ACE19, the collagen binding subdomain of Enterococus faecalis surface protein ACE. Proteins *69*, 199-203.

Ponnuraj,K., Xu,Y., Moore,D., Deivanayagam,C.C., Boque,L., Hook,M., and Narayana,S.V. (2002). Crystallization and preliminary X-ray crystallographic analysis of Ace: a collagen-binding MSCRAMM from Enterococcus faecalis. Biochim. Biophys. Acta *1596*, 173-176.

Popa,M.P., McKelvey,T.A., Hempel,J., and Hendrix,R.W. (1991). Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. J. Virol. *65*, 3227-3237.

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

Price,N.C., Dwek,R.A., Ratcliffe,R.G., and Wormald,M. (2001). Principles and Problems in Physical Chemistry for Biochemists. Oxford University Press.

Qiagen. QIAexpress(R) Detection and Assay Handbook. 3, 1-103. 2002.
Ref Type: Pamphlet

Richards,F.M. (1958). On The Enzymic Activity of Subtilisin-Modified Ribonuclease. Proc. Natl. Acad. Sci. U. S. A *44*, 162-166.

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

Schmidt,T.G., Koepke,J., Frank,R., and Skerra,A. (1996). Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. J. Mol. Biol. *255*, 753-766.

Schwarz-Linek, U. 2011.
Ref Type: Personal Communication

Shiba,K. and Schimmel,P. (1992). Functional assembly of a randomly cleaved protein. Proc. Natl. Acad. Sci. U. S. A *89*, 1880-1884.

Singh,A., Mai,D., Kumar,A., and Steyn,A.J. (2006). Dissecting virulence pathways of Mycobacterium tuberculosis through protein-protein association. Proc. Natl. Acad. Sci. U. S. A *103*, 11346-11351.

Smith,G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315-1317.

Spraggon,G., Koesema,E., Scarselli,M., Malito,E., Biagini,M., Norais,N., Emolo,C., Barocchi,M.A., Giusti,F., Hilleringmann,M., Rappuoli,R., Lesley,S., Covacci,A., Masignani,V., and Ferlenghi,I. (2010). Supramolecular organization of the repetitive backbone unit of the Streptococcus pneumoniae pilus. PLoS. One. *5*, e10919.

Striebel,F., Imkamp,F., Sutter,M., Steiner,M., Mamedov,A., and Weber-Ban,E. (2009). Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. Nat Struct. Mol. Biol.

Stroppolo,M.E., Falconi,M., Caccuri,A.M., and Desideri,A. (2001). Superefficient enzymes. Cell Mol. Life Sci. *58*, 1451-1460.

Telford,J.L., Barocchi,M.A., Margarit,I., Rappuoli,R., and Grandi,G. (2006). Pili in gram-positive pathogens. Nat. Rev. Microbiol. *4*, 509-519.

Terpe,K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. *60*, 523-533.

Tesar,M., Beckmann,C., Rottgen,P., Haase,B., Faude,U., and Timmis,K.N. (1995). Monoclonal antibody against pIII of filamentous phage: an immunological tool to study pIII fusion protein expression in phage display systems. Immunotechnology. *1*, 53-64.

Thermo Scientific. SPDP Crosslinkers. 2008. USA.
Ref Type: Pamphlet

van de Rijn,I. and Kessler,R.E. (1980). Growth characteristics of group A streptococci in a new chemically defined medium. Infect. Immun. *27*, 444-448.

Vengadesan,K., Ma,X., Dwivedi,P., Ton-That,H., and Narayana,S.V. (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.

Voet,D. and Voet,J.G. (2004). Biochemistry: Biomolecules, Mechanisms of Enzyme Action, and Metabolism. John Wiley & Sons Inc.).

Voss,S. and Skerra,A. (1997). Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. Protein Eng *10*, 975-982.

Ward,D.T., Brown,E.M., and Harris,H.W. (1998). Disulfide bonds in the extracellular calcium-polyvalent cation-sensing receptor correlate with dimer formation and its response to divalent cations in vitro. J. Biol. Chem. *273*, 14476-14483.

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.

Wilken,J. and Kent,S.B. (1998). Chemical protein synthesis. Curr. Opin. Biotechnol. *9*, 412-426.

Xu,M.Q. and Evans,T.C., Jr. (2005). Recent advances in protein splicing: manipulating proteins in vitro and in vivo. Curr. Opin. Biotechnol. *16*, 440-446.

Yamamoto,T. and Nagamune,T. (2009). Expansion of the sortase-mediated labeling method for site-specific N-terminal labeling of cell surface proteins on living cells. Chem. Commun. (Camb.) 1022-1024.

Yeates,T.O. and Clubb,R.T. (2007). How some pili pull. Science *318*, 1558-1559.

Yoshihara,H.A., Mahrus,S., and Wells,J.A. (2008). Tags for labeling protein N-termini with subtiligase for proteomics. Bioorg. Med. Chem. Lett. *18*, 6000-6003.

Zahner,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.

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

Zong,Y., Xu,Y., Liang,X., Keene,D.R., Hook,A., Gurusiddappa,S., Hook,M., and Narayana,S.V. (2005). A 'Collagen Hug' model for Staphylococcus aureus CNA binding to collagen. EMBO J. *24*, 4224-4236.

Appendix

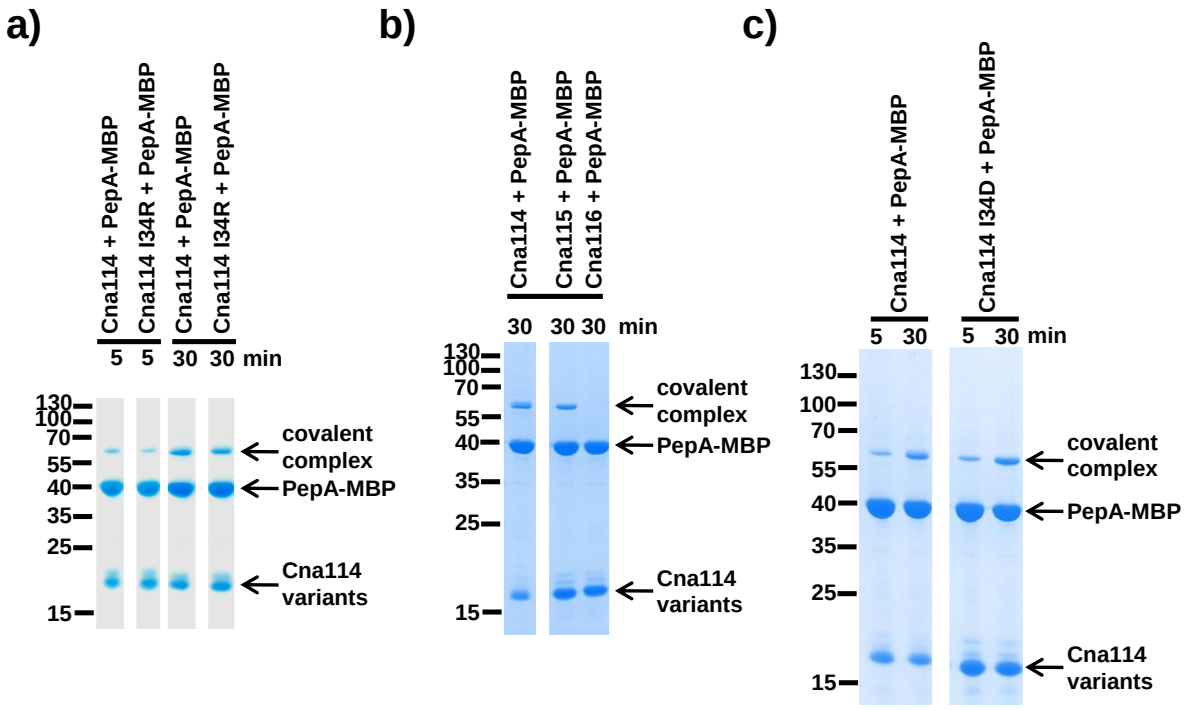


Fig. A.1. Reconstitution of Cna114 mutants. a) Reconstitution of Cna114 compared to Cna114 I34R with PepA-MBP. Proteins were mixed at 10 µM in PBS pH 7.5 and incubated at 25°C and samples were removed at 5 and 30 min. They were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining. The lanes shown were all part of the same experiment but were cropped to show relevant lanes with a white space in between sections that were cropped. b) Reconstitution of Cna114 compared to Cna115 and Cna116 with PepA-MBP. Proteins were mixed at 10 µM in PBS pH 7.5 and incubated at 25°C for 30 min. Samples were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. The lanes shown were all part of the same experiment but were cropped to show relevant lanes with a white space in between sections that were cropped. c) Reconstitution of Cna114 compared to Cna114 I34D with PepA-MBP. Proteins were mixed at 10 µM in PBS pH 7.5 and incubated at 25°C and samples were removed at 5 and 30 min. They were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 10% SDS-PAGE with Coomassie staining. The lanes shown were all part of the same experiment but were cropped to show relevant lanes with a white space in between sections that were cropped.

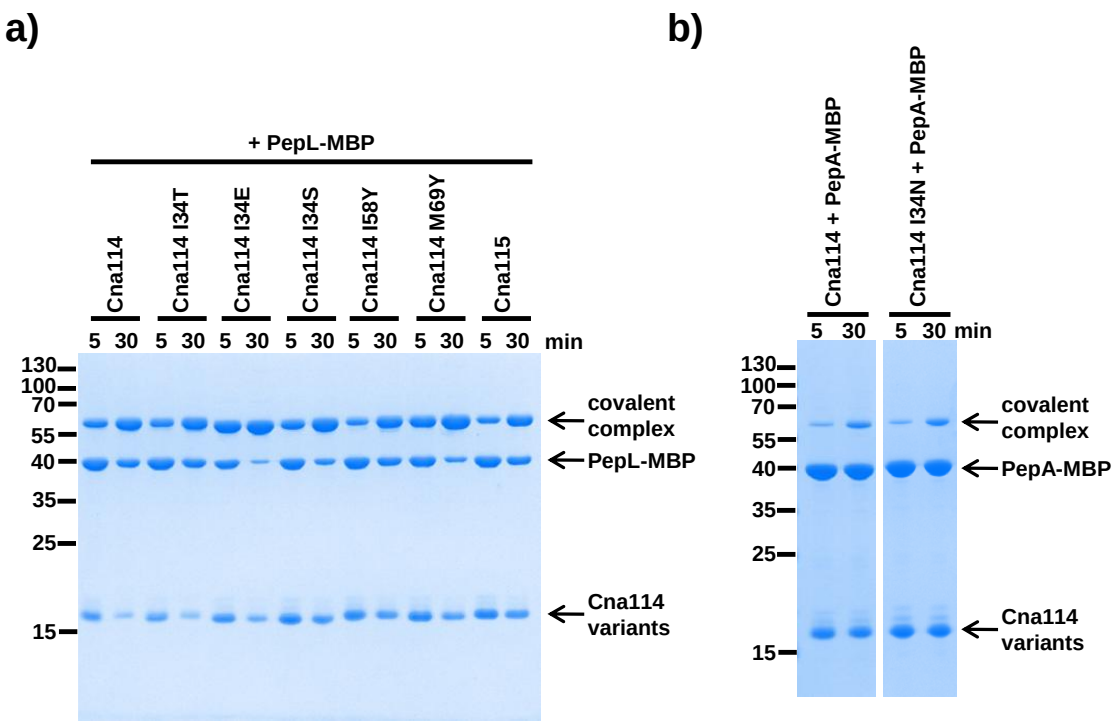


Fig. A.2. Reconstitution of Cna114 variants. a) Reconstitution of Cna114 compared to Cna114 I34T, Cna114 I34E, Cna114 I34S, Cna114 I58Y, Cna114 M69Y, and Cna115 with PepL-MBP. Proteins were mixed at 10 µM in PBS pH 7.5 and incubated at 25°C and samples were removed at 5 and 30 min. They were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. b) Reconstitution of Cna114 compared to Cna114 I34N with PepA-MBP. Proteins were mixed at 10 µM in PBS pH 7.5 and incubated at 25°C and samples were removed at 5 and 30 min. They were then heated to 95°C for 7 min in SDS loading buffer and analyzed on 14% SDS-PAGE with Coomassie staining. The lanes shown were all part of the same experiment but were cropped to show relevant lanes with a white space in between sections that were cropped.

The following article has been removed from this version of the thesis for copyright reasons: Zakeri, B. & Howarth, M. (2010), 'Spontaneous intermolecular amide bond formation between side chains for irreversible peptide targeting', Journal of the American Chemical Society 132, 4526-4527. Available at http://dx.doi.org/10.1021/ja910795a

Spontaneous Intermolecular Amide Bond Formation between Side Chains for Irreversible Peptide Targeting

Bijan Zakeri and Mark Howarth*

Department of Biochemistry, Oxford University, South Parks Road, Oxford, OX1 3QU, UK.
E-mail: mark.howarth@bioch.ox.ac.uk

Supporting figures

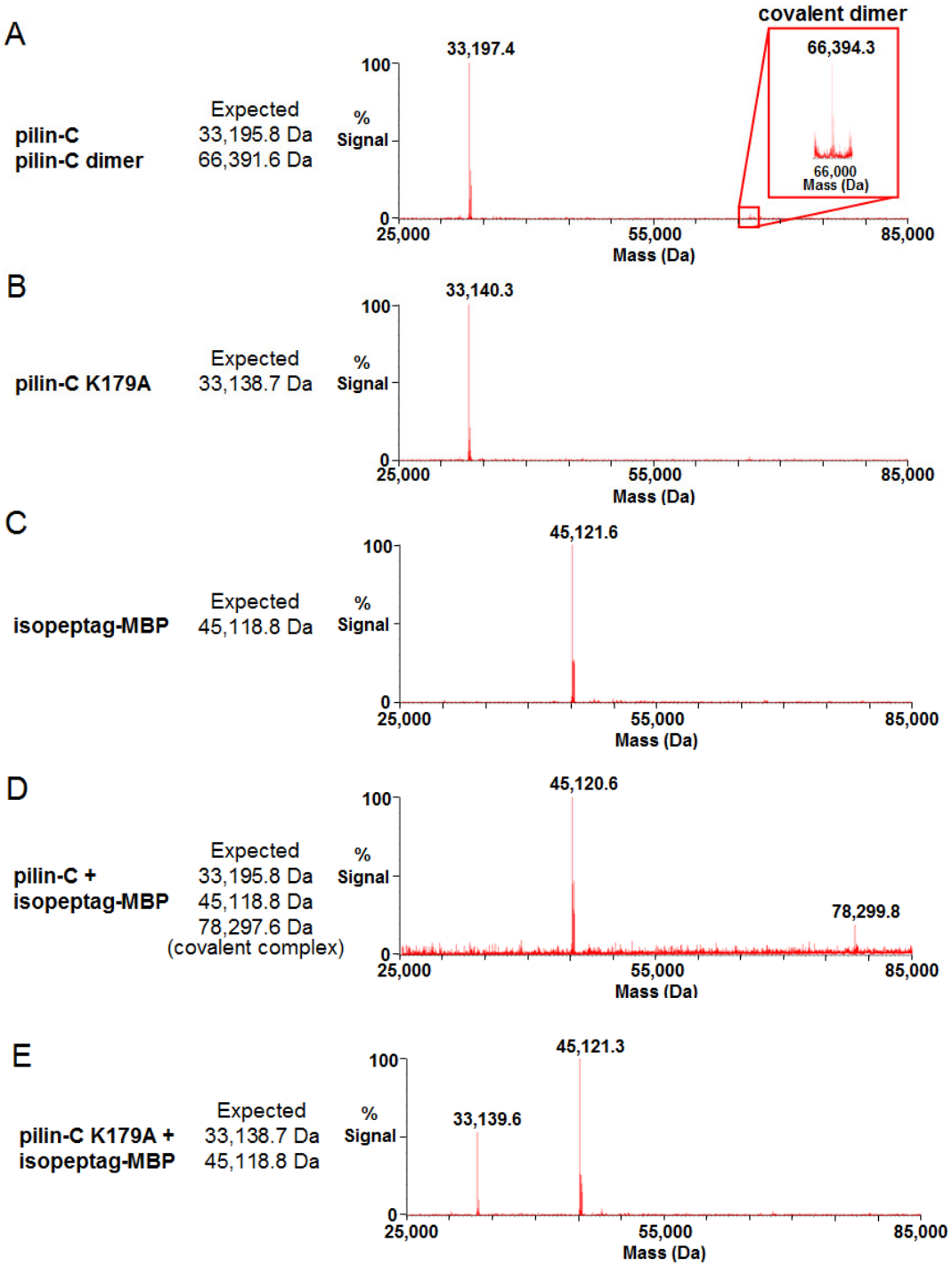


Figure S1. Mass spectrometry of spontaneous intermolecular amide bond formation. (A) Pilin-C alone. Expression of pilin-C also yields a small amount of covalent pilin-C dimer, visible here by MS (inset) and also visible by SDS-PAGE after boiling (Figure 1D). Pilin-C contains one isopeptide in the N-domain. Based on its mass (loss of 17 Da for each isopeptide), the pilin-C dimer contains two isopeptides, one intramolecular and one intermolecular. (B) Pilin-C K179A alone, the negative control for intermolecular amide bond formation. (C) Isopeptag-MBP alone. (D) Pilin-C and isopeptag-MBP were mixed at 100 μM for 16 h at 4 ºC in PBS. The covalent complex was observed, corresponding to loss of NH_3 (17 Da) upon amide bond formation, as well as some residual isopeptag-MBP. (E) A negative control, where pilin-C K179A and isopeptag-MBP were mixed as in (D) but no high molecular weight complex was detected.

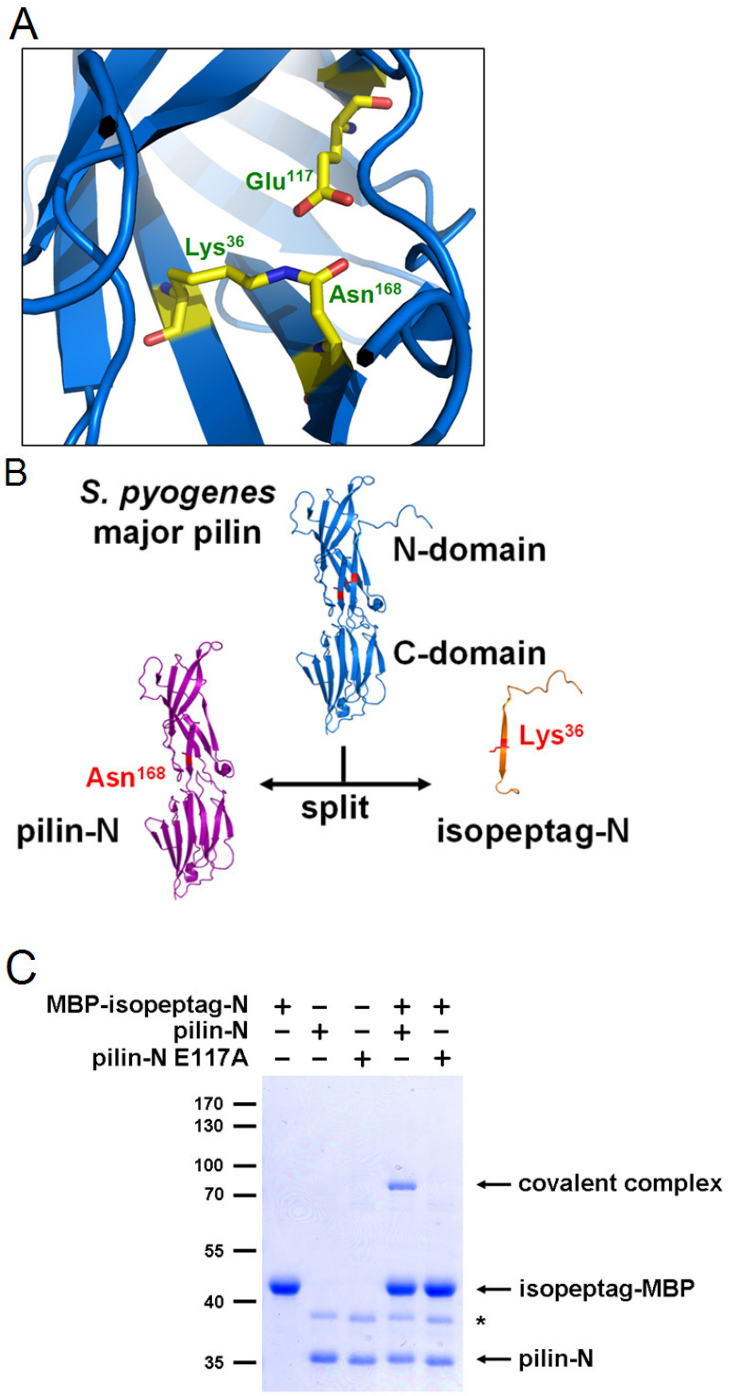


Figure S2. An alternative spontaneous intermolecular amide bond-forming peptide. (A) Key residues for spontaneous amide bond formation in the N-domain of the major pilin protein (Spy0128) of *S. pyogenes*, from Protein Data Bank 3B2M. (B) Isopeptag-N construction. Spy0128 was dissected into a small N-terminal fragment (isopeptag-N in orange) and a large C-terminal fragment (pilin-N in purple). Amide bond forming residues are highlighted in red. (C) Isopeptag-N and pilin-N were able to covalently react *in vitro*. MBP-isopeptag-N and pilin-N, each at 10 μM, were mixed and incubated for 24 h at 25 ˚C, pH 7.0. Pilin-N E117A is a negative control, where the Glu essential for promoting covalent reaction was mutated.[1] Reaction was analyzed by SDS-PAGE and Coomassie staining. *A side-product of pilin-N expression, likely from misformation of the intramolecular isopeptide.
Sequence of isopeptag-N: ATTVHGETVVNGAKLTVTKNLDLVNSNA.

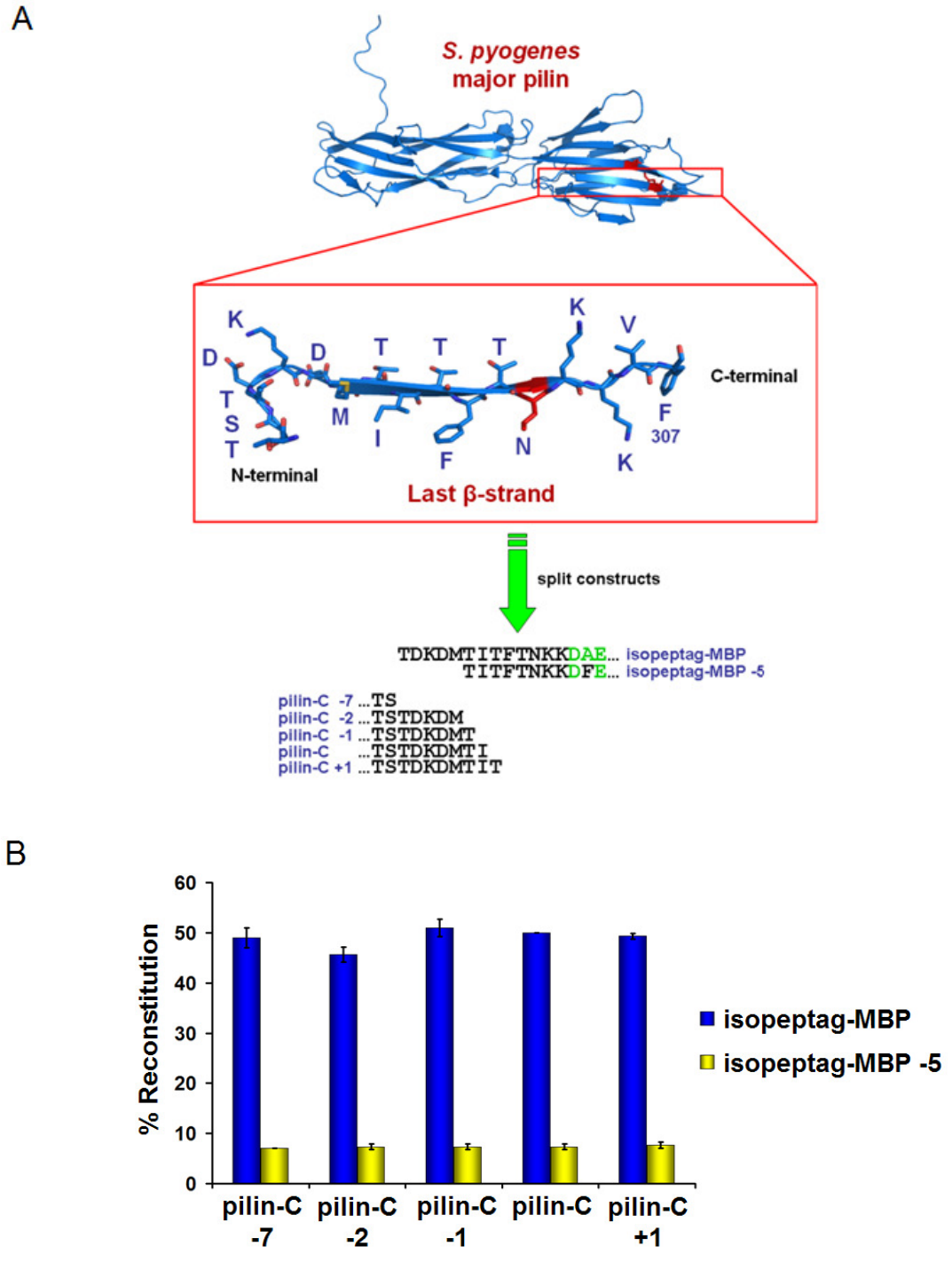


Figure S3. Optimization of split pilin constructs to generate isopeptag and pilin-C. (A) Schematic illustrating the various isopeptag-MBP and pilin-C constructs tested. The reactive Asn is shown in red. The Spy0128 crystal structure (Protein Data Bank 3B2M) stops at residue F307 and has Val at 306 instead of the Asp in *S. pyogenes* M1 GAS (NCBI Entrez Nucleotide accession no. AE004092). Our isopeptag construct had Asp at residue 306 and we also changed the surface-exposed Phe at position 307 to Ala, to increase solubility. Residues different to the crystal structure are shown in green. The numbers depict the change in the number of residues compared to the pilin-C and isopeptag-MBP constructs finally chosen and used in all other figures. Including in the protein partner residues from the final β-strand was evaluated on the basis that these residues could inhibit the other two neighboring β-strands interacting with each other and so occluding the isopeptag binding site. (B) Reconstitution efficiency of the isopeptag-MBP and pilin-C constructs, each at 10 μM, after incubation at 25 ºC for 24 h in PBS pH 7.4, determined by SDS-PAGE and Coomassie staining. The means of triplicate results are shown ± 1 s.d.

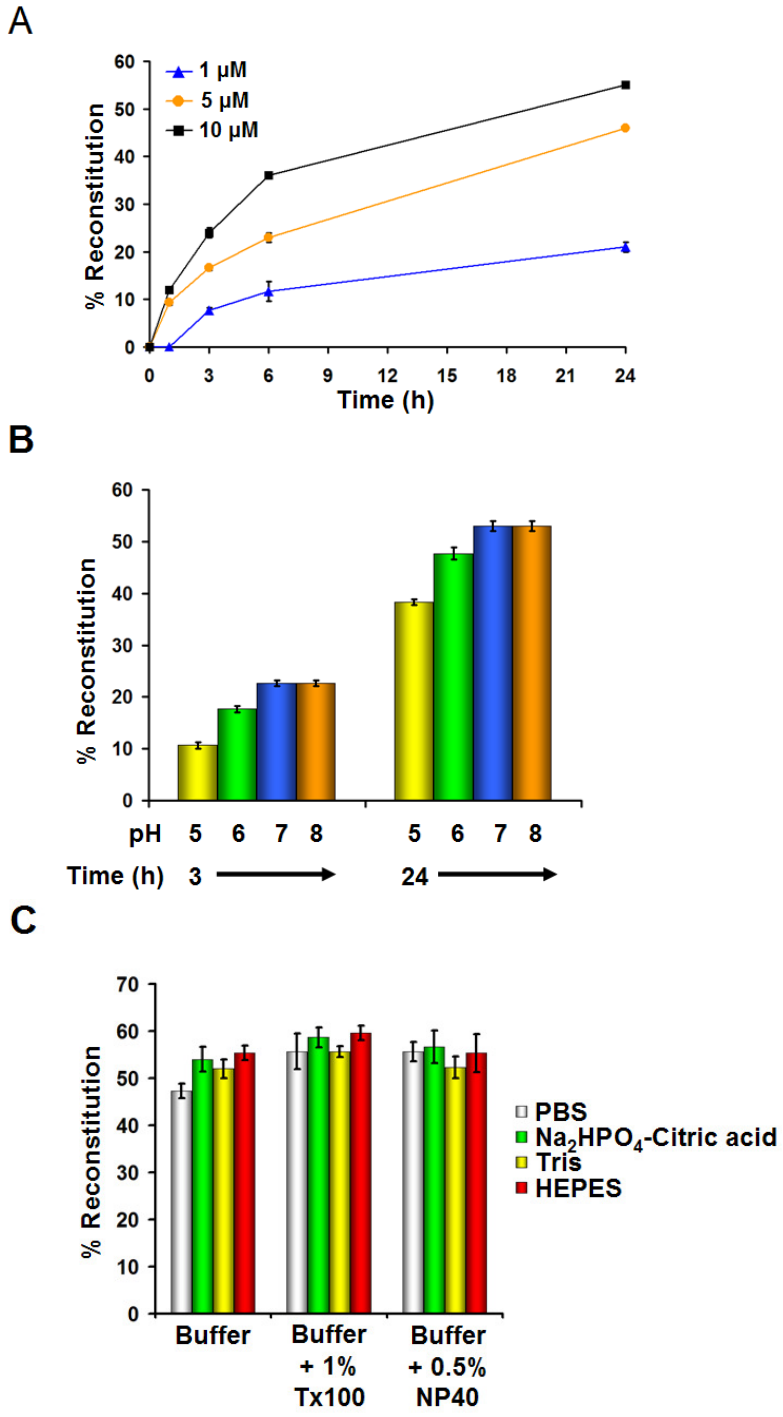


Figure S4. Factors affecting spontaneous amide bond formation. (A) Concentration-dependence of spontaneous amide bond formation. Isopeptag-MBP and pilin-C were mixed each at 1, 5, or 10 µM and incubated at 25 ºC, pH 7.0. The reaction was stopped at varying times and analyzed by SDS-PAGE with Coomassie staining. The means of triplicate results are shown ± 1 s.d. (some errors bars are too small to be visible). (B) pH-dependence of spontaneous amide bond formation. Isopeptag-MBP and pilin-C were mixed at 10 µM in 40 mM Na_2HPO_4 - 20 mM citric acid at the indicated pH and incubated at 25 ºC. The reaction was stopped at varying times and analyzed by SDS-PAGE with Coomassie staining. The means of triplicate results are shown ± 1 s.d. (C) Buffer-dependence of spontaneous amide bond formation. Isopeptag-MBP and pilin-C each at 10 µM were mixed in the indicated buffer with or without the common biological detergents Triton X-100 (Tx100) or Nonidet P-40 (NP40) and incubated for 24 h at 25 ºC, pH 7.0. Detergent might have been expected to change the rate of reconstitution, if amide bond formation depended on exposure of a hydrophobic part of pilin-C. PBS is phosphate buffered saline. Reaction was analyzed by SDS-PAGE with Coomassie staining. The means of triplicate results are shown ± 1 s.d.

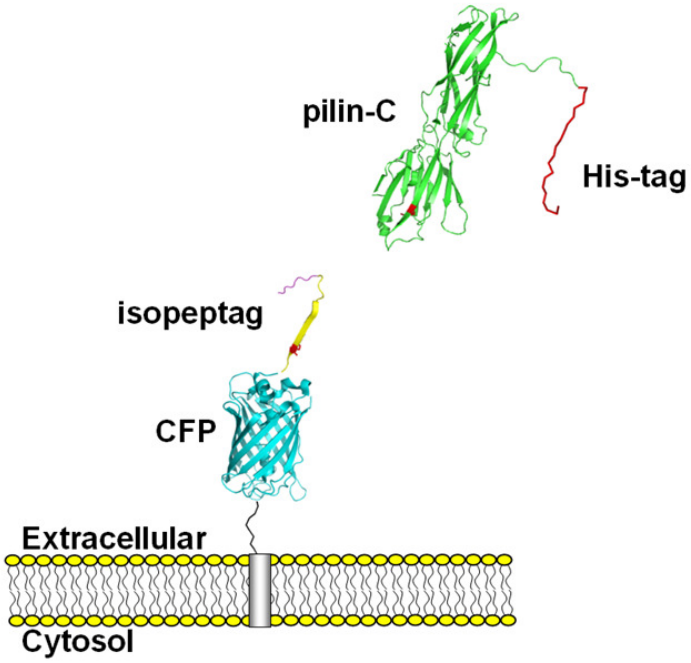


Figure S5. Cartoon of the isopeptag-CFP-TM construct, used to target the isopeptag to the surface of mammalian cells where it could be detected by exogenous pilin-C. CFP = cyan fluorescent protein. TM = transmembrane helix.

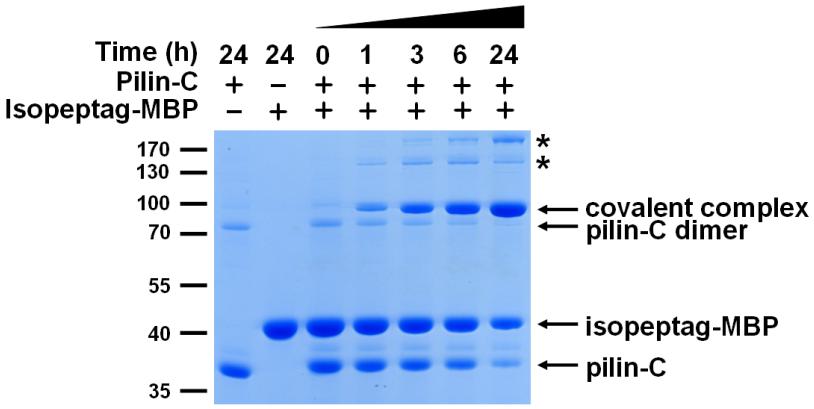


Figure S6. Time-course of spontaneous amide bond formation by SDS-PAGE. Isopeptag-MBP and pilin-C were mixed at 10 μM and incubated at 25 ºC, pH 7.0. The reaction was stopped at varying times and analyzed by SDS-PAGE with Coomassie staining. *This side-product likely relates to the pilin-C dimer reacting with 1 or 2 isopeptag-MBP.

Supporting methods

Plasmids

All residue numbers are based on the Spy0128 protein.[1] PCR was performed with KOD Hot Start DNA polymerase (Roche) and cycling conditions as recommended by the manufacturer. The *Spy0128* gene was PCR-amplified from *Streptococcus pyogenes* M1 GAS strain SF370 genomic DNA (ATCC 700294D-5) using the primers 5'–GGGGCATATGGCTACAACAGTTCACGGG–3' and 5'–GGGGAAGCTTTTATTCAAAGTCTTTTTTATTTG–3', and subsequently cloned in to pET28a (Novagen) using NdeI and HindIII. To construct isopeptag-MBP, isopeptag (Spy0128 residues 293-308) was PCR-amplified from pET28a-Spy0128 using the primers 5'–GGGGCATATGGGAACTGATAAAGATATGACC–3' and 5'–ACCACTTTCACCACTACCTTCAAAGTCTTTTTTATTTG–3', while the maltose binding protein (MBP) was amplified from pMAL (New England Biolabs) using the primers 5'–AAGGTAGTGGTGAAAGTGGTAAAATCGAAGAAGGTAAA–3' and 5'–GGGGAAGCTTTTACGAGCTCGAATTAGTCTG–3'. The two fragments were joined by overlap-extension PCR, and cloned in to pET28a using *NdeI* and *HindIII* sites. This gave a GSGESG linker between isopeptag and MBP. The final isopeptag-MBP construct was generated by a further mutation of F307A (based on the complete *Spy0128* sequence), to remove an exposed hydrophobic residue,[1] using the QuikChange™ (Stratagene) protocol with KOD Hot Start DNA polymerase and the primers 5'– TTTACAAATAAAAAAGACGCTGAAGGTAGTGGTGAAAG–3' and its reverse complement. To generate isopeptag-MBP -5 (Spy0128 residues 298-308), the sequence was PCR-amplified from pET28a-isopeptag-MBP using primers 5'–GGGGCATATGGGAACCATTACTTTTACAAAT–3' and 5'–GGGGAAGCTTTTACGAGCTCGAATTAGTCTG–3' and cloned in to pET28a using *NdeI* and *HindIII* sites. To generate pilin-C (Spy0128 residues 18-299), a stop codon was introduced at residue 300 in pET28a-Spy0128 using the primer 5'–ACAAGAGACATCTACTGATAAAGATATGACCATTTAGTTTACAAATAAAAAAGACTTTGAATAAAAGCTTG–3' and its reverse complement. To generate pilin-C K179A from pET28a-Spy0128, we used the primer 5'–TCTACTACATTAACGGTGAAGGCAAAAGTTTCAGGTACCGGTGG–3' and its reverse complement, and then a stop codon was introduced at residue 300 in pET28a-Spy0128 using the primer 5'–ACAAGAGACATCTACTGATAAAGATATGACCATTTAGTTTACAAATAAAAAAGACTTTGAATAAAAGCTTG–3' and its reverse complement. To generate pilin-C -7 (Spy0128 residues 18-292), a stop codon was introduced at residue 293 in pET28a-Spy0128 using the primer 5'–GCAGGTAATTCAACTGAACAAGAGACATCTTAGGATAAAGATATGACCATTACTTTTACAAAT–3' and its reverse complement. To generate pilin-C -2 (Spy0128 residues 18-297), a stop codon was introduced at residue 298 in pET28a-Spy0128 using the primer 5'–ACTGAACAAGAGACATCTACTGATAAAGATATGTAGATTACTTTTACAAATAAAAAAGACTTTGAAGTG–3' and its reverse complement. To generate pilin-C -1 (Spy0128 residues 18-298), a stop codon was introduced at residue 299 in pET28a-Spy0128 using the primer 5'–GAACAAGAGACATCTACTGATAAAGATATGACCTAGACTTTTACAAATAA

AAAAGACTTTGAATAAAAG–3' and its reverse complement. To generate pilin-C +1 (Spy0128 residues 18-300), a stop codon was introduced at residue 301 in pET28a-Spy0128 using the primer 5'–CAAGAGACATCTACTGATAAAGATATGACCATTACTTAGACAAATAAAAAAGACTTTGAATAAAA–3' and its reverse complement.

To construct MBP-isopeptag-N (Figure S2B), isopeptag-N (Spy0128 residues 18-45) was PCR-amplified from pET28a-Spy0128 using the primers 5'–CTAATTCGAGCTCGGGTTCGGGTGAAAGTGGTGCTACAACAGTTCACGGG–3' and 5'–GGGGAAGCTTTTATGCATTGCTATTAACTAAATC–3', while MBP was amplified from pMAL using primers 5'–CAAGCATATGAAAATCGAAGAAG–3' and 5'–CGAACCCGAGCTCGAATTAGTCTG–3'. The two fragments were joined using overlap-extension PCR, and cloned in to pET28a at *NdeI* and *HindIII* sites. Pilin-N (Spy0128 residues 46-308) was generated from pET28a-Spy0128 using the primers 5'–GGGGCATATGGGATTAATTCCAAATACAGAT–3' and 5'–GGGGAAGCTTCTAGTGATGGTGATGGTGATGTCCTGATCCTTCAAAGTCTTTTTTATTTG–3', and subsequently cloned in to pET28a using *NdeI* and *HindIII* sites. Pilin-N E117A was generated from pET28a-pilin-N using the primer 5'–GTGTTTATTATTACAAAGTAACTGCGGAGAAGATAGATAAAGTTCCTGG–3' and its reverse complement. Isopeptag-CFP-TM was based on pDisplay (Invitrogen). 5'–GACAGATCTGGCGGCACTGATAAAGATATGACCATTACTTTTACAAATAAAAAAGACTTTGAAGGTAGTGGTATGGTGAGCAAGGGCGAG–3' and 5'–ACTCTCGGCATGGACGAGCTATACAAGCGGCCGCGGGGAG–3' were used for PCR from AP-CFP-TM.[2] The PCR product was gel purified and inserted in the *BglII* and *SacII* sites of pDisplay.

To construct a bicistronic vector for simultaneous expression of pilin-C or pilin-C K179A and isopeptag-MBP (Figure 3A), the pilin-C and pilin-C K179A constructs were PCR-amplified from pET28a-pilin-C or pET28a-pilin-C K179A, using primers 5'– GGGGAAGCTTCTCTAGAAATAATTTTGTTTAAC–3' and 5'–GGGGCTCGAGCTAAATGGTCATATCTTTATC–3'. The PCR products were gel purified and inserted in to the *HindIII* and *XhoI* sites of pET28a-isopeptag-MBP, to give pET28a-(isopeptag-MBP + pilin C) and pET28a-(isopeptag-MBP + pilin C K179A). From this plasmid, from a single T7 promoter, an mRNA would be transcribed containing first a sequence that coded for the isopeptag-MBP protein, then a spacer following the isopeptag-MBP stop codon, then a Shine-Dalgarno sequence to promote translation, and finally a sequence coding for the pilin C or pilin-C K179A protein.

We verified all constructs and mutations by sequencing. The nuclear co-transfection marker pECFP-H2B (human histone H2B for nuclear localization fused to enhanced cyan fluorescent protein) was constructed as described.[3]

Protein expression

All proteins were expressed using *E. coli* BL21 DE3 RIPL cells (Stratagene), grown in LB with 0.8 % glucose and 0.05 mg/mL kanamycin. We diluted overnight cultures 100-fold, grew at 37 ºC to A_{600} 0.5, and induced with 0.4 mM IPTG for 4 h at 30 °C. All proteins were purified by nickel affinity chromatography, using standard methods, and dialyzed into PBS. Protein concentration was determined from A_{280}, using the extinction coefficient predicted by ExPASy ProtParam. Typical expression yields per L of culture were 10 mg for pilin-C and 16 mg for isopeptag-MBP.

SDS-PAGE

SDS-PAGE was performed on 10% polyacrylamide gels, using an X-cel SureLock (Invitrogen) at 200 V. Samples were heated at 95 °C for 7 min in SDS loading buffer on a Bio-Rad C1000 thermal cycler before loading. Gels were stained with Coomassie brilliant blue, destained in 60% MilliQ water, 30% methanol and 10% acetic acid, and band intensities were quantified using a ChemiDoc XRS imager and QuantityOne 4.6 software (Bio-Rad).

Reconstitution reactions

Reactions for analyzing speed, temperature-dependence, and pH-dependence of amide bond formation by SDS-PAGE were performed with 10 μM of each protein in 40 mM Na_2HPO_4 - 20 mM citric acid at the indicated pH. For Figure 2C, one reactant was maintained at 10 μM while the concentration of the other reactant was varied. For analyzing amide bond formation in various buffers and detergents (Figure S4C), 10 μM of pilin-C and 10 μM of isopeptag-MBP were mixed in either phosphate buffered saline (PBS) pH 7.4, 40 mM Na_2HPO_4 - 20 mM citric acid pH 7.4, 50 mM Tris (tris-hydroxymethyl aminomethane) pH 7.4, or 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazine ethanesulfonate) pH 7.4, and also in the presence and absence of 1% Triton X-100 or 0.5% Nonidet P-40. pH for all buffers was adjusted with 1 M NaOH or 1 M HCl, except for HEPES which was adjusted with 1 M KOH to establish whether Na^+ was required for reaction. All reactions were incubated for 24 h unless otherwise stated. Reactions were stopped by adding SDS loading buffer and heating at 95 ºC for 7 min. For 4 ºC reactions, samples were incubated in the refrigerator; for 25 ºC and 37 ºC reactions, samples were incubated on a Bio-Rad C1000 thermal cycler with a heated lid to prevent evaporation. The percentage reconstitution between pilin-C and isopeptag-MBP was calculated by dividing the density of the band for the pilin-C:isopeptag-MBP covalent complex by the sum of the density of all the bands in the lane, then multiplying by 100. To calculate the percentage of pilin-C reacted (Figure 2C), the percentage reduction in intensity of the pilin-C band was determined relative to the controls where no reaction partner was added. The percentage of isopeptag-MBP that reacted was calculated by the same approach.

For testing *in situ* reconstitution within living cells (Figure 3A), *E. coli* expressing monocistronic or bicistronic plasmids were induced for 4 h at 30 ºC, as above. Samples of bacterial cultures were mixed with SDS loading buffer and immediately heated at 95 °C for 7 min to prevent further reconstitution, before analysis by SDS-PAGE.

Mass spectrometry (Figure 2A and S1)

We performed mass spectrometry with a Micromass LCT time-of-flight electrospray ionization mass spectrometer (Micromass UK). 100 μM of each protein was mixed in PBS pH 7.4 and incubated at 4 ºC for 16 h. The reactions were then dialyzed against MilliQ H_2O using 0.025 μm VSWP Millipore membrane filters (VSWP 01300) for 1 h at 25 ºC. The reactions were then further de-salted using Millipore ZipTip pipette tips (ZTC04S096). The m/z spectrum was converted to a molecular mass profile using Maximum Entropy processing by the software

MassLynx V4.00.00. Predicted masses were determined from the amino acid sequence by ExPASy ProtParam, subtracting 17 Da for every isopeptide bond.

Cell culture, labeling and microscopy (Figure 3B)

HeLa cells were grown in DMEM with 10% Fetal Calf Serum, 50 U/mL penicillin, and 50 μg/mL streptomycin. Cells were transfected using 0.75 μL Lipofectamine 2000 (Invitrogen) and 0.3 μg of isopeptag-CFP-TM per well of a 48-well plate, according to manufacturer's instructions. One day after transfection, cells were incubated with 5 μM of either pilin-C or pilin-C K179A and incubated at 37 ºC for 4 h. Cells were then stained at 4 ºC as follows: wash 3 times with PBS + 5mM $MgCl_2$ (PBS-Mg), add 100 μL of 11 μg/mL penta-His antibody (Qiagen) in PBS-Mg with 1% BSA and incubate for 10 min, wash 3 times with PBS-Mg, add 100 μL of 20 μg/mL Alexa Fluor 555-anti-mouse antibody (Invitrogen) in PBS-Mg with 1% BSA and incubate for 10 min, wash 3 times with PBS-Mg. Cells were fixed with 100 μL fixing solution (PBS containing 4% formaldehyde, 4% sucrose, and 5 mM $MgCl_2$) for 15 min at 4 ºC and then washed 3 times with PBS-Mg. Cells were imaged using a wide-field DeltaVision Core fluorescent microscope (AppliedPrecision) with a 40× oil-immersion lens. ECFP (436DF20 excitation, 480DF40 emission, Chroma 86002v1 dichroic) and Alexa Fluor 555 (540D420 excitation, 600DF50 emission, Chroma 84100bs polychroic) images were collected and analyzed using softWoRx 3.6.2 software. Typical exposure times were 0.1-1.0 s and fluorescence images were background-corrected. Different samples in the same experiment were prepared, imaged and analyzed under identical conditions.

References

(1) Kang, H. J.; Coulibaly, F.; Clow, F.; Proft, T.; Baker, E. N. *Science* **2007**, *318*, 1625.

(2) Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. *Nat. Methods* **2005**, *2*, 99.

(3) Platani, M.; Goldberg, I.; Lamond, A. I.; Swedlow, J. R. *Nat. Cell Biol.* **2002**, *4*, 502.