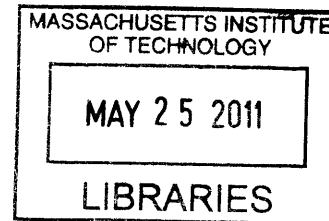


# Site-specific Protein Labeling via Sortase A and its Applications

by

Maximilian Wei-Lin Popp

B.S. Molecular Genetics  
B.A. Chemistry  
University of Rochester, 2005



**ARCHIVES**

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY

AT THE  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

© 2011 Massachusetts Institute of Technology  
All rights reserved

Signature of author: .....  
Maximilian Wei-Lin Popp  
May 16, 2011

Certified by: .....  
Hidde L. Ploegh  
Professor of Biology  
Thesis supervisor

Accepted by: .....  
Robert Sauer  
Professor of Biology  
Co-chair of the Biology Graduate Committee

# **Site-specific Protein Labeling via Sortase A and its Applications**

by

**Maximilian Wei-Lin Popp**

**Submitted to the Department of Biology  
on May 16, 2011 in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy in  
Biology**

## **Abstract**

Technological improvements in the assays and equipment used for biological, biochemical, biophysical and microscopy purposes have ensured that methods for labeling of proteins with reporter molecules remain in high demand. Standard chemical labeling methods using entities that react with amino acid side chains lack the specificity to ensure precise placement of reporter groups. Genetic methods, although specific, lack the versatility afforded by chemical synthesis-most reporters are limited to protein sized domains or peptide tags to which corresponding antibody based reagents are available.

The first portion of this work is devoted to the establishment of a system that allows for the site-specific labeling of proteins with a wide variety of chemically synthesized probes. This system exploits sortases, a class of bacterial transpeptidases, that recognize a small five amino acid tag genetically fused to the protein of interest and catalyze the formation of an amide bond between the protein to be studied and the probe.

The second part of this thesis describes how this sortase mediated protein labeling method has been implemented to explore enzyme structure and function, improve the physical properties of therapeutic proteins, study glycoproteins important for innate immune responses in living cells (Appendix A), and visualize influenza glycoproteins in living, infected cells.

Finally, a protocol is included for this system (Appendix B), which is both versatile and easy to establish in any lab. The synthetic chemistry demanded is minimal, requiring only standard, readily available reagents, making the system amenable to labs equipped for cell and molecular biology experiments.

Thesis Supervisor: Hidde L. Ploegh  
Title: Professor of Biology



<b>Title Page</b> .....	<b>1</b>
<b>Abstract</b> .....	<b>2</b>
<b>Table of Contents</b> .....	<b>3</b>
<b>Acknowledgements</b> .....	<b>7</b>
<b>Chapter 1</b> .....	<b>8</b>
<b>Introduction- Making and Breaking Peptide Bonds: Protein engineering using sortase</b>	
Introduction.....	9
Biological function and Biochemistry of Sortase A .....	9
Bacterial Surface Engineering .....	11
C-terminal Labeling .....	12
N-terminal Labeling.....	14
Labeling at N- and C-termini.....	15
Post-translational Modification Mimics .....	16
Piecemeal Assembly of Proteins/Protein Domains/Peptides.....	17
Anchoring to Solid Surfaces .....	18
Protein Expression and Purification.....	19
Conclusion .....	20
References.....	21
Figure Legends.....	27
Figures.....	28
Tables.....	31

**Part I: Development of Sortase-mediated transpeptidation technology**

<b>Chapter 2</b> .....	<b>34</b>
<b>Sortagging: A novel method for site-specific labeling of proteins</b>	
Abstract .....	35
Main Text.....	35
Methods.....	42
References.....	47
Figure Legends.....	49
Figures.....	52
Supplementary Figure Legends .....	57
Supplementary Figures .....	61
Supplementary Text.....	71
Print Article.....	82

**Chapter 3 ..... 85**  
**Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities**

Abstract.....	86
Main Text.....	86
References.....	92
Figure Legends.....	93
Figures.....	94
Supplementary Figure Legends .....	96
Supplementary Figures .....	98
Supplementary Text.....	103
Print Article.....	121

**Chapter 4 ..... 124**  
**A straight path to circular proteins**

Abstract.....	125
Introduction.....	125
Results.....	127
Discussion.....	134
Methods.....	138
References.....	144
Figure Legends.....	147
Figures.....	151
Supplementary Figure Legends .....	157
Supplementary Figures .....	161
Supplementary Text.....	170
Print Article.....	172

**Part II. Applications of the sortase methodology**

**Chapter 5 ..... 194**  
**Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations**

Abstract.....	195
Introduction.....	195
Results.....	198
Discussion.....	204
Methods.....	207
References.....	213
Figure Legends.....	214
Figures.....	220
Supplementary Figure Legends .....	228

Supplementary Figures .....	230
Print Article.....	233
<b>Chapter 6 .....</b>	<b>243</b>
<b>Sortase-catalyzed transformations that improve the properties of cytokines</b>	
Abstract.....	244
Introduction.....	244
Results.....	246
Discussion.....	251
Methods.....	253
References.....	258
Figure Legends.....	260
Figures.....	265
Supplementary Figure Legends .....	270
Supplementary Figures .....	272
Supplementary Text.....	277
Print Article.....	282
<b>Chapter 7 .....</b>	<b>289</b>
<b>A system for monitoring influenza glycoproteins in living infected cells</b>	
Abstract.....	290
Introduction.....	290
Results.....	292
Discussion.....	296
Methods.....	296
References.....	300
Figure Legends.....	301
Figures.....	306
<b>Chapter 8 .....</b>	<b>314</b>
<b>Future Directions</b>	
Text .....	315
References.....	321
Figure Legends.....	322
Figures.....	323
Print article (for Chapter 1 and Chapter 8) .....	325
German version of print article (for Chapter 1 and Chapter 8).....	334

**Appendix A.....345**  
**Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages**

Abstract..... 346  
Introduction..... 346  
Results..... 349  
Discussion..... 357  
Methods..... 361  
References..... 371  
Figure Legends..... 374  
Figures..... 380

**Appendix B.....387**  
**Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation**

Introduction..... 388  
Expression and Purification of Sortase A ..... 388  
Design of Protein Substrates for Sortase-mediated Transpeptidation ..... 391  
Design of Peptide Probes Compatible with Transpeptidation ..... 392  
Labeling Purified Proteins ..... 393  
Labeling Cell-surface Proteins in Living Cells..... 394  
Background Information..... 396  
Critical Parameters..... 398  
Troubleshooting ..... 399  
Anticipated Results ..... 400  
Time Considerations ..... 400  
References..... 401  
Figure Legends..... 403  
Print Article..... 404

## **Acknowledgements**

First, I'm indebted to Hidde for letting me work on this thesis in his lab. At every step along the way he has given me opportunities that I would not have had elsewhere, and I'm forever grateful for this. I've enjoyed all of our discussions, especially the ones peppered with crazy ideas from out of left-field. I've realized that the ideas that initially seem outlandish are usually the ones that are the most fun to roll around in my brain during my long T-ride home.

I'm also indebted to many people in the Ploegh lab: Gijs Grotenbreg, for getting me up and running; Howard Hang for making me really, really enjoy talking about science, reading papers, and thinking broadly; John Antos and Nick Yoder, for constant advice, collaboration, and chemistry hand-holding; Carla, Van der Veen and Ludo, Kat, and Elizabeth for interesting chats and diversions (related to science and, more importantly, unrelated to science); and the rest of the members of the lab for always offering interesting ideas to think about.

I thank my parents for supporting me but never pushing me and for instilling in me an interest in science. The German version of the review from Chapter 8 would not have been possible without my Dad. I also thank my mother- and father-in-law for their support and for letting me steal Andrea away to Boston.

Lastly, I thank Andrea for supporting me through graduate school with equal parts of bringing-home-the-bacon and emotional support. I'm very grateful for the past six years of love and patience and for making it so much fun to not be working in lab, with all of the little trips, meals, and diversions. Also, thanks for coming in every weekend with me when I needed to do work, even if it was only for the Clear Flour treats afterwards.

# **Chapter 1: Introduction**

## Chapter 1: Introduction

**Making and Breaking Peptide Bonds: Protein engineering using sortase**  
(Adapted from: Making and Breaking Peptide Bonds: Protein engineering using sortase, Angewandte Chemie, Accepted)

### **Introduction**

Sortases are a class of bacterial enzymes that possess transpeptidase activity. It is the ability to break a peptide bond site-specifically and then reform a new bond with an incoming nucleophile that makes sortase an attractive tool for protein engineering. This technique has been adopted for a range of applications running the gamut from chemistry-based to cell biological to technological. We provide a brief overview of the biology of the sortase enzyme and current applications to protein engineering. We identify areas that lend themselves to further innovation and that suggest new applications.

### **Biological Function and Biochemistry of Sortase A**

Many gram-positive bacteria display virulence factors on their cell wall for successful colonization and pathogenesis<sup>1</sup>. Anchoring of proteins to the bacterial cell wall is the purview of sortase enzymes<sup>2 3 4</sup>, a class of thiol-containing transpeptidases. These enzymes recognize substrate proteins bearing a “sorting motif” (LPXTG in the case of *Staphylococcus aureus*) and harbor a catalytic cysteine used to cleave the peptide bond between the threonine and glycine within this pentapeptide<sup>5 6 7</sup>. Other sortases from different bacterial species use the same or similar recognition sequences<sup>8 9 10</sup>. For a database of sortase and their substrates, see [http://bamics3.cmbi.kun.nl/cgi-bin/jos/sortase\\_substrates/index.py](http://bamics3.cmbi.kun.nl/cgi-bin/jos/sortase_substrates/index.py)<sup>11 12</sup>. This reaction initially yields a thioacyl intermediate<sup>13 14</sup>, in a fashion analogous to the mechanism used

## Chapter 1: Introduction

by cysteine proteases. Where cysteine proteases use water to resolve this intermediate and generate a hydrolysis product, sortase will accept the N-terminus of an oligo-glycine nucleophile, resulting in the creation of a new peptide bond (**Figure 1.1**). In the course of the normal sorting reaction, the pentaglycine crossbridge in the lipid-II cell wall precursor carries out the nucleophilic attack on the acyl-enzyme<sup>15</sup>. The cell wall precursor with its covalently attached protein is then incorporated into the growing peptidoglycan layer.

In addition to anchoring virulence factors to the cell wall, sortases build the pilus structure that many bacteria use for attachment to host cells and to form biofilms<sup>16 17 18 19</sup>. The details of this process differ between bacterial species<sup>20 21 22 23 24</sup>, but in general terms, it involves a sortase that polymerizes pilin subunits bearing both a sorting signal and a nucleophilic  $\epsilon$ -amine of a lysine in an internally located motif<sup>25</sup>. This protein-protein ligation reaction results in polymerization of the pilin subunits, but does not mediate anchoring of the growing polymer to the cell wall. This is the job of the housekeeping sortase, which accepts the lipid-II precursor nucleophile<sup>26</sup>.

Sortases represent a bona-fide drug target because of their central role in virulence<sup>27 28 29 30 31</sup>. For this reason they have been studied extensively both structurally<sup>32-34 32 34 35</sup> and biochemically<sup>36 37 38</sup>. The structure of sortase A from *S. aureus* consists of an eight stranded beta-barrel-like fold, termed the sortase fold, with a hydrophobic cleft formed by the  $\beta 7$  and  $\beta 8$  strands, surrounded by the  $\beta 3$ – $\beta 4$ ,  $\beta 2$ – $\beta 3$ ,  $\beta 6$ – $\beta 7$ , and  $\beta 7$ – $\beta 8$  loops (**Figure 1.2**). This cleft houses the catalytic cysteine residue (Cys 184) and accommodates



## Chapter 1: Introduction

substrate binding. An additional structural feature of the *S. aureus* enzyme is a calcium binding site formed by the  $\beta 3/\beta 4$  loop. Calcium binding to this site coordinates a residue in the  $\beta 6/\beta 7$  loop, slowing its motion, allowing substrate to bind, and stimulating activity by eight-fold<sup>39</sup>. The biochemical details of the active site include a key histidine residue (H120) that can form a thiolate-imidazolium ion pair with the catalytic cysteine<sup>40</sup>. It is the deprotonated form of cysteine that is competent for catalysis. However, at physiological pH, the ionized forms of these key amino acids are in equilibrium with the neutral forms, and only a small percentage ( $\sim 0.06\%$ ) of the total enzyme is catalytically competent at any given time<sup>41-42</sup>. The cysteine attacks the amide bond between the threonine and glycine residues in the sorting motif. The protonated imidazolium ion acts as a general acid for the departing glycine  $\alpha\text{NH}_2$ , and gives rise to an acylated form of sortase. An incoming glycine nucleophile is then deprotonated, attacks the thioester, and re-establishes an amide bond. If instead water attacks the acyl-enzyme intermediate, the reaction yields the dead-end hydrolysis product<sup>43</sup>.

### **Bacterial Surface Engineering**

The sortase-mediated system of anchoring proteins to the cell wall of gram-positive bacteria was first exploited to decorate these microbes with heterologous proteins. Such experiments require the creation of a genetic fusion of the heterologous protein to the sorting motif. The heterologous protein is then expressed and directed to the surface through the normal cell wall sorting pathway. In this manner, the enzyme alkaline phosphatase has been anchored to the cell wall of *Staphylococcus aureus*<sup>44</sup>, the E7 protein of HPV16 has been displayed on *S. gordonii*, a commensal microbe in the oral

## Chapter 1: Introduction

cavity<sup>45</sup>, and alpha amalyase has been affixed to the peptidoglycan of *B. subtilis*, helped by co-expression of the sortase gene from *L. monocytogenes*<sup>46</sup>. The peptidoglycan cell wall can even be decorated with non-natural entities (fluorescein, biotin, azide) by incubation of dividing *S. aureus* cultures with chemical probes appended to the N-terminus of an LPXTG peptide<sup>47</sup>. Incorporation of what are in essence N-terminal labeling probes (see later section) uses the endogenous sortase enzyme and anchors the exogenously provided probes onto available pentaglycine sidechains of the cell wall.

### **C-terminal Labeling**

The ability of sortase to recognize the sorting motif when transplanted onto recombinantly expressed proteins allows the site-specific incorporation of moieties and functional groups that cannot be encoded genetically (**Figure 1.1**). This method requires only that the LPXTG motif be solvent exposed and usually results in high yields of the desired transpeptidation product. Indeed, many substrate proteins have now been labeled with probes bearing a wide range of functionalities, including biotin, fluorophores, cross-linkers and multi-functional probes (**Table 1.1, Table 1.2**)<sup>48</sup>. Labeling of recombinant proteins by sortase A requires no sophisticated synthetic chemistry; most of the probes are readily accessible by standard peptide synthesis, using off-the shelf reagents.

Production and folding of recombinant substrate proteins is not usually compromised by the presence of the small LPXTG tag. Because all transformations are carried out using sortase under physiological buffer conditions (pH, ionic strength, ionic requirements) on substrates whose proper folding and activity status can be ascertained prior to starting the reaction, loss of biological activity is rarely, if ever, observed for the final product. The

## Chapter 1: Introduction

ability to engage in a sortase-catalyzed transacylation appears to be determined solely by accessibility and flexibility of the sorting motif. Intein-based protein engineering methods usually require that substrates first fold while fused to a protein-sized intein domain, at times causing solubility issues<sup>49</sup>.

The utility of the sortase labeling method stems from the fact that the enzyme tolerates substrates unrelated in structure and sequence immediately upstream from the cleavage site. This property is not unexpected, given the role of sortase in anchoring a broad range of protein substrates to the cell wall. The substrate need not even be proteinaceous-peptide-nucleic acids linked to the sortase cleavage site can be ligated to a glycine-linked cell penetrating peptide (MAP model amphipathic peptide) to yield active antisense PNA-CPP conjugates<sup>50</sup>. Likewise, the identity of the substituents C-terminal to the glycine nucleophile seem to matter not at all: D-amino acid-containing peptides, folate, branched protein transduction domains<sup>51</sup>, and large polyethylene glycol chains<sup>52</sup> have all been attached using sortase. The cleavage site need not even be near the C-terminus of the substrate protein. A sufficiently large solvent-exposed loop will suffice. This property has been exploited to investigate the contribution to substrate binding and catalysis of a key loop in the deubiquitinating enzyme UCHL3<sup>53</sup>. Because the cleavage site can be placed in a loop, it is possible to interrupt the connectivity of the protein backbone, while simultaneously installing a reporter moiety (biotin or fluorophore) to monitor the behavior of the cleaved enzyme in the presence of uncleaved, wild-type enzyme. This trait is likely to apply to many proteins whose conformation includes an exposed, flexible loop.

## Chapter 1: Introduction

The C-terminal labeling technique is particularly useful for the study of type-II membrane proteins embedded in the living mammalian cell membrane. Type II membrane proteins have C-termini that are exposed to extracellular space and thus are excellent candidates for sortase-mediated labeling. Proteins with this type II topology have been particularly refractory to genetic fusion with fluorescent proteins. N-terminal placement of a fluorescent protein usually impedes cotranslational insertion of the type II membrane protein into the ER, while C-terminal tagging with GFP places this bulky substituent close to the site of interaction with ligands of the type II membrane protein in question. . CD40L, influenza neuraminidase<sup>54</sup>, and osteoclast differentiation factor (ODF)<sup>55</sup> have all been labeled in live cells in this way.

### **N-terminal Labeling**

Protein labeling at the N-terminus can be accomplished simply by moving the placement of the sortase recognition element from the protein to the short peptide probe and by inclusion of a suitable number of glycines at the target protein's N-terminus (**Figure 1.1**). Both methyl-ester mimetics of the sortase motif<sup>56</sup> as well as the complete LPXTG sortase recognition motif can be used as scaffolds for such probes<sup>57</sup>. Conceptually, this labeling technique is analogous to the C-terminal labeling, except the acyl-enzyme intermediate is generated between sortase and the peptide probe, and the protein to be labeled bears several glycines at the N-terminus, the  $\alpha\text{NH}_2$  of which serves as the nucleophile. This strategy was used to install fluorescent probes at the N-terminus of membrane proteins in living mammalian cells after a clever initial unmasking step by

## Chapter 1: Introduction

sortase itself to expose the nucleophilic glycine<sup>57</sup>. This system was later used to install reporter fluorophores on the N-terminus of the G protein-coupled receptor, PAFR. This allowed the direct observation of the trafficking of the cell surface-exposed pool after labeling. PAFR receptors with key mutations were then shown to traffic aberrantly<sup>58</sup>. For both the C-terminal and N-terminal labeling of cell surface proteins, the sortase technique allows access only to the cell surface pool of the protein of interest. This is an advantage when interested in exploring the behavior of only the surface exposed fraction of a particular protein. If ligand binding is restricted to the cell surface, then this is also usually the relevant fraction. Genetic fusions to fluorescent proteins, by their very nature report on the protein of interest from the moment of its genesis inside the cell and onwards. Although this trait comes with its own advantages, it may complicate the distinction between proteins in the course of their biosynthesis and the behavior of the mature, biologically relevant pool of protein. The sortase-based strategies should thus be viewed as a useful adjunct to the GFP-based methods, but with the added benefit of increased chemical flexibility.

### **Labeling at N- and C-termini**

By using sortases with distinct substrate specificity, it is possible to combine N-terminal and C-terminal labeling strategies. The *Streptococcus pyogenes* enzyme<sup>59</sup> (SrtA<sub>Strep</sub>) recognizes and cleaves the LPXTA motif and accepts alanine-based nucleophiles. It also cleaves the SrtA<sub>Staph</sub> LPXTG motif, albeit with reduced efficiency. In contrast, the SrtA<sub>Staph</sub> enzyme does not cleave the LPXTA motif, and thus the two enzymes are orthogonal with respect to the LPXTA sequence. This property was exploited to label

## Chapter 1: Introduction

both termini of GFP and UCHL3 with unique fluorophores. A masking strategy was used in which the N-terminal glycines needed for SrtA<sub>Staph</sub> labeling were exposed after proteolytic cleavage by thrombin. This step avoids protein oligomerization, likely to occur during the SrtA<sub>Strep</sub> C-terminal labeling step<sup>56</sup>.

### **Post-translational Modification Mimics**

Sortase methodology allows the production of homogenous recombinant protein preparations that are modified with non-genetically templated post-translational modifications. Glycoproteins, normally elaborated by a complex set of enzymatic events in the secretory pathway, can thus be constructed. LPXTG-tagged proteins and peptides can be modified with 6-aminohexose-based sugar nucleophiles, including aminoglycoside antibiotics and their analogs<sup>60</sup>. Glycosylphosphatidylinositol (GPI) anchors, normally attached at the C-terminus of proteins can be phenocopied by ligation of LPXTG peptides to synthetic glycine nucleophiles, in turn linked to the phosphoethanolamine moiety on a GPI derivative<sup>61</sup>.

Various peptides (CD52 fragment, MUC1 peptide) and small proteins (CD24) have been attached to GPI mimics with trisaccharide cores<sup>62-63</sup><sup>63</sup>. Lipidation of proteins is yet another important post-translational modification that is poorly studied because of the lack of tools available to obtain homogenous preparations of lipoproteins. Sortase has been used to fill this void<sup>64</sup>. A glycine-based scaffold was modified with a panel of linear alkyl chains (length C12 - C24) as well as with cholesterol or amantadine, and then used to modify a suitably LPETG-tagged version of eGFP. These eGFP lipoproteins associated with the plasma membranes of living cells in a chain-length dependent

## Chapter 1: Introduction

fashion, the optimum being C22, from where they gained access to the endosomal compartment.

### **Piecemeal Assembly of Proteins/Protein Domains/Peptides**

Folded proteins with an exposed glycine at the N-terminus may serve as nucleophiles for sortase labeling. Substrate proteins bearing the LPXTG motif can be fused to the incoming nucleophile protein, creating large transpeptidation products. By using independently folded proteins as substrates, it is possible to avoid many of the solubility issues that plague expression of large genetically encoded fusions. This property was exploited to facilitate NMR structural analysis, which typically requires highly concentrated protein preparations, making poor solubility a major obstacle. Sortase was used to attach an unlabeled, and hence NMR invisible, solubility enhancing tag (G<sub>3</sub>-GB1) onto the C-terminus of the Vav SH3 domain, a domain that is nearly insoluble by itself at neutral pH<sup>65</sup>. The structure of the attached <sup>13</sup>C/<sup>15</sup>N-labeled Vav SH3 domain was then resolved by NMR, without confounding signals from the solubility-enhancing tag. A versatile panel of immunodetection reagents has been created using sortase. Protein-protein ligations were carried out between an Fc binding module (ZZ-domain) and several detection enzymes (AP, Luc, GOD) using sortase<sup>66</sup>. Mucin glycopeptides were synthesized that contain both N- and O-linked glycans with the help of sortase as a synthetic tool. Separate peptides bearing either O- or N- linked glycans were constructed by a combination of chemical synthesis and elaboration of the glycan structure by enzymatic synthesis. These glycopeptides were then stitched together using sortase to yield a stereochemically homogenous preparation<sup>67 68</sup>. Sortase has been applied to the

## Chapter 1: Introduction

construction of G-protein coupled receptor (GPCR) mimics through a combined recombinant, enzymatic, and chemical synthesis (CRECS) strategy<sup>69</sup>. GPCRs are 7-pass transmembrane proteins that use three extracellular loops as well as the extracellular N-terminal segment to bind their ligands. To mimic this arrangement, three loops were made synthetically, cyclized by native chemical ligation, and appended to a triglycine linked peptide scaffold. Then the GPCR N-terminus, fused to the sortase cleavage site, was recombinantly expressed in *E.coli* and attached to the scaffold via sortase-mediated ligation. These elegantly engineered soluble mimics should allow the systematic characterization of the contributions of each region to ligand binding and represent a true marriage between what can be accomplished through chemical synthesis and molecular biology.

A unique variant of the protein-protein ligation occurs when the LPXTG motif and N-terminal glycines are both present in the same construct. If both units are sufficiently close in space in the folded protein, the N-terminus can form a peptide bond with the sortase recognition element, resulting in a stable, circular transpeptidation product<sup>70 52</sup>. Circular proteins have useful biochemical properties. They are resistant to aggregation, require more energy for denaturation, and since they lack exposed termini, are resistant in their native form to exoprotease attack<sup>71 72 73 74 75</sup>.

### **Anchoring to Solid Surfaces**

Covalent immobilization of proteins onto solid supports has been accomplished by sortase. By relying on the enzyme specificity to anchor substrates to the solid surface,



## Chapter 1: Introduction

proteins are immobilized uniformly and in defined orientation for subsequent exposure to the analyte of interest, a major advantage of the method. Stringent wash conditions can be employed because of the stable amide bond that links the protein to the surface, as was done by covalently attaching GFP to glycine-derivatized polystyrene beads<sup>52</sup>.

Attachment of adhesion proteins from gram-positive bacteria to fluorescent glycine-derivatized polystyrene beads was done in a similar manner<sup>76</sup>. Anchoring of GFP and Tus proteins to glycidyl methacrylate beads derivatized with oligo glycine, as well as to glycine-modified agarose resin (Affi-Gel), and glycine modified aminosilane coated glass slides has been achieved<sup>77</sup>. Directional anchoring of proteins onto triglycine-modified carboxymethylated dextran-based Biacore sensor chips for use in surface plasmon resonance has also been accomplished<sup>78</sup>. Recombinant fibronectin-binding protein (rFba-LPETG) from group A streptococcus (GAS) was anchored in this manner, which then allowed the measurement of binding of human factor H to the immobilized protein. With an aim to develop the reagents needed for chemoenzymatic synthesis of glycoconjugates, immobilized  $\beta$ 1,4-galactosyltransferase (rhGalT) and *Helicobacter pylori*  $\alpha$ 1,3-fucosyltransferase (rHFucT) were covalently attached to alkylamine-sepharose beads. These enzymes are both active and reusable when directionally anchored to the solid-phase<sup>79</sup>.

### **Protein Expression and Purification**

Genetic fusions between sortase and a protein of interest have been constructed for the purposes of protein purification. A linear fusion between hexahistidine-tagged sortase, followed by the LPETG tag, and the protein of interest is first purified by Ni-NTA IMAC

## Chapter 1: Introduction

and then cleaved off of the resin by addition of  $\text{Ca}^{+2}$  and triglycine to yield highly pure protein with one additional glycine<sup>80</sup> (**Fig. 1.3**). This method was adapted for protein production in a wheat germ cell-free translation system, with the goal of creating a general purification method that can be used in automated, high throughput protein production. In this version, a biotin acceptor peptide (a 15 amino acid peptide that is the target of *E. coli* biotin ligase) replaces the hexahistidine tag and the proteins are purified with strepavidin resin in the presence of calcium chelators to prevent premature cleavage<sup>81</sup>. Both methods yield the target protein with one extra glycine at the N-terminus, a configuration that is poised for N-terminal labeling by sortase if desired (**Figure 1.3**). Sortase A from *S. aureus* is an extremely soluble enzyme that can be produced in high yield (>40mg/L of culture)-this property has been exploited to enhance the solubility of proteins of interest by fusion to a version of sortase lacking the catalytic cysteine<sup>82</sup>.

### **Conclusion**

This thesis will describe the efforts we have made to establish the sortase-mediated labeling system for N-terminal, C-terminal and dual-terminus labeling of proteins. We have expanded the utility of the method to include labeling of purified proteins, proteins in heterogenous complex mixtures, and on the living cell surface. A diverse array of functionalities has been appended to proteins in each case, creating a toolbox that has expanded what can be accomplished by marrying molecular biology with chemical synthesis. We further describe how we have applied this technology to accomplish feats

## Chapter 1: Introduction

that would not have been possible otherwise in the areas of protein engineering and cell biology.

### References

1. Marraffini, L.A., Dedent, A.C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* **70**, 192-221 (2006).
2. Mazmanian, S.K., Liu, G., Ton-That, H. & Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760-763 (1999).
3. Mazmanian, S.K., Ton-That, H. & Schneewind, O. Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Mol Microbiol* **40**, 1049-1057 (2001).
4. Zink, S.D. & Burns, D.L. Importance of srtA and srtB for growth of Bacillus anthracis in macrophages. *Infect Immun* **73**, 5222-5228 (2005).
5. Ton-That, H. & Schneewind, O. Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction. *J Biol Chem* **274**, 24316-24320 (1999).
6. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc Natl Acad Sci U S A* **96**, 12424-12429 (1999).
7. Kruger, R.G., *et al.* Analysis of the substrate specificity of the Staphylococcus aureus sortase transpeptidase SrtA. *Biochemistry* **43**, 1541-1551 (2004).
8. Dramsi, S., Trieu-Cuot, P. & Bierne, H. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. *Res Microbiol* **156**, 289-297 (2005).
9. Pucciarelli, M.G., *et al.* Identification of substrates of the Listeria monocytogenes sortases A and B by a non-gel proteomic analysis. *Proteomics* **5**, 4808-4817 (2005).
10. Comfort, D. & Clubb, R.T. A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect Immun* **72**, 2710-2722 (2004).
11. Boekhorst, J., de Been, M.W., Kleerebezem, M. & Siezen, R.J. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J Bacteriol* **187**, 4928-4934 (2005).
12. Pallen, M.J., Lam, A.C., Antonio, M. & Dunbar, K. An embarrassment of sortases - a richness of substrates? *Trends Microbiol* **9**, 97-102 (2001).
13. Ton-That, H., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J Biol Chem* **275**, 9876-9881 (2000).

## Chapter 1: Introduction

14. Aulabaugh, A., *et al.* Development of an HPLC assay for Staphylococcus aureus sortase: evidence for the formation of the kinetically competent acyl enzyme intermediate. *Anal Biochem* **360**, 14-22 (2007).
15. Perry, A.M., Ton-That, H., Mazmanian, S.K. & Schneewind, O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J Biol Chem* **277**, 16241-16248 (2002).
16. Ton-That, H. & Schneewind, O. Assembly of pili in Gram-positive bacteria. *Trends Microbiol* **12**, 228-234 (2004).
17. Scott, J.R. & Zahner, D. Pili with strong attachments: Gram-positive bacteria do it differently. *Mol Microbiol* **62**, 320-330 (2006).
18. Mandlik, A., Swierczynski, A., Das, A. & Ton-That, H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol* **16**, 33-40 (2008).
19. Proft, T. & Baker, E.N. Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* **66**, 613-635 (2009).
20. Ton-That, H. & Schneewind, O. Assembly of pili on the surface of Corynebacterium diphtheriae. *Mol Microbiol* **50**, 1429-1438 (2003).
21. Ton-That, H., Marraffini, L.A. & Schneewind, O. Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae. *Mol Microbiol* **53**, 251-261 (2004).
22. Manzano, C., *et al.* Sortase-mediated pilus fiber biogenesis in Streptococcus pneumoniae. *Structure* **16**, 1838-1848 (2008).
23. Neiers, F., *et al.* Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J Mol Biol* **393**, 704-716 (2009).
24. Manzano, C., Izore, T., Job, V., Di Guilmi, A.M. & Dessen, A. Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens. *Biochemistry* **48**, 10549-10557 (2009).
25. Budzik, J.M., *et al.* Amide bonds assemble pili on the surface of bacilli. *Proc Natl Acad Sci U S A* **105**, 10215-10220 (2008).
26. Mandlik, A., Das, A. & Ton-That, H. The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proc Natl Acad Sci U S A* **105**, 14147-14152 (2008).
27. Mazmanian, S.K., Liu, G., Jensen, E.R., Lenoy, E. & Schneewind, O. Staphylococcus aureus sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc Natl Acad Sci U S A* **97**, 5510-5515 (2000).
28. Paterson, G.K. & Mitchell, T.J. The biology of Gram-positive sortase enzymes. *Trends Microbiol* **12**, 89-95 (2004).
29. Maresso, A.W. & Schneewind, O. Sortase as a target of anti-infective therapy. *Pharmacol Rev* **60**, 128-141 (2008).
30. Suree, N., Jung, M.E. & Clubb, R.T. Recent advances towards new anti-infective agents that inhibit cell surface protein anchoring in Staphylococcus aureus and other gram-positive pathogens. *Mini Rev Med Chem* **7**, 991-1000 (2007).

## Chapter 1: Introduction

31. Paterson, G.K. & Mitchell, T.J. The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. *Microbes Infect* **8**, 145-153 (2006).
32. Ilangovan, U., Iwahara, J., Ton-That, H., Schneewind, O. & Clubb, R.T. Assignment of the <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N signals of Sortase. *J Biomol NMR* **19**, 379-380 (2001).
33. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O. & Clubb, R.T. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* **98**, 6056-6061 (2001).
34. Zong, Y., Bice, T.W., Ton-That, H., Schneewind, O. & Narayana, S.V. Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J Biol Chem* **279**, 31383-31389 (2004).
35. Suree, N., *et al.* The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J Biol Chem* **284**, 24465-24477 (2009).
36. Kruger, R.G., Dostal, P. & McCafferty, D.G. Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal Biochem* **326**, 42-48 (2004).
37. Marraffini, L.A., Ton-That, H., Zong, Y., Narayana, S.V. & Schneewind, O. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. A conserved arginine residue is required for efficient catalysis of sortase A. *J Biol Chem* **279**, 37763-37770 (2004).
38. Frankel, B.A., Tong, Y., Bentley, M.L., Fitzgerald, M.C. & McCafferty, D.G. Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* **46**, 7269-7278 (2007).
39. Naik, M.T., *et al.* *Staphylococcus aureus* Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J Biol Chem* **281**, 1817-1826 (2006).
40. Ton-That, H., Mazmanian, S.K., Alksne, L. & Schneewind, O. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J Biol Chem* **277**, 7447-7452 (2002).
41. Connolly, K.M., *et al.* Sortase from *Staphylococcus aureus* does not contain a thiolate-imidazolium ion pair in its active site. *J Biol Chem* **278**, 34061-34065 (2003).
42. Frankel, B.A., Kruger, R.G., Robinson, D.E., Kelleher, N.L. & McCafferty, D.G. *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* **44**, 11188-11200 (2005).
43. Huang, X., *et al.* Kinetic mechanism of *Staphylococcus aureus* sortase SrtA. *Biochemistry* **42**, 11307-11315 (2003).
44. Schneewind, O., Model, P. & Fischetti, V.A. Sorting of protein A to the staphylococcal cell wall. *Cell* **70**, 267-281 (1992).
45. Pozzi, G., *et al.* Delivery and expression of a heterologous antigen on the surface of streptococci. *Infect Immun* **60**, 1902-1907 (1992).

## Chapter 1: Introduction

46. Nguyen, H.D. & Schumann, W. Establishment of an experimental system allowing immobilization of proteins on the surface of *Bacillus subtilis* cells. *J Biotechnol* **122**, 473-482 (2006).
47. Nelson, J.W., *et al.* A Biosynthetic Strategy for Re-engineering the *Staphylococcus aureus* Cell Wall with Non-native Small Molecules. *ACS Chem Biol* (2010).
48. Popp, M.W., Antos, J.M. & Ploegh, H.L. Site-specific protein labeling via sortase-mediated transpeptidation. *Curr Protoc Protein Sci* **Chapter 15**, Unit 15 13 (2009).
49. Vila-Perello, M. & Muir, T.W. Biological applications of protein splicing. *Cell* **143**, 191-200 (2010).
50. Pritz, S., *et al.* Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. *J Org Chem* **72**, 3909-3912 (2007).
51. Mao, H., Hart, S.A., Schink, A. & Pollok, B.A. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **126**, 2670-2671 (2004).
52. Parthasarathy, R., Subramanian, S. & Boder, E.T. Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation. *Bioconjug Chem* **18**, 469-476 (2007).
53. Popp, M.W., Artavanis-Tsakonas, K. & Ploegh, H.L. Substrate filtering by the active site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations. *J Biol Chem* **284**, 3593-3602 (2009).
54. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**, 707-708 (2007).
55. Tanaka, T., Yamamoto, T., Tsukiji, S. & Nagamune, T. Site-specific protein modification on living cells catalyzed by Sortase. *Chembiochem* **9**, 802-807 (2008).
56. Antos, J.M., *et al.* Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* **131**, 10800-10801 (2009).
57. Yamamoto, T. & Nagamune, T. 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 (2009).
58. Hirota, N., *et al.* Amino acid residues critical for endoplasmic reticulum export and trafficking of platelet-activating factor receptor. *J Biol Chem* **285**, 5931-5940 (2010).
59. Race, P.R., *et al.* Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J Biol Chem* **284**, 6924-6933 (2009).
60. Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V.K. & Roy, R.P. Peptide-sugar ligation catalyzed by transpeptidase sortase: a facile approach to neoglycoconjugate synthesis. *J Am Chem Soc* **130**, 2132-2133 (2008).
61. Guo, X., Wang, Q., Swarts, B.M. & Guo, Z. Sortase-catalyzed peptide-glycosylphosphatidylinositol analogue ligation. *J Am Chem Soc* **131**, 9878-9879 (2009).
62. Wu, Z., Guo, X., Wang, Q., Swarts, B.M. & Guo, Z. Sortase A-catalyzed transpeptidation of glycosylphosphatidylinositol derivatives for chemoenzymatic synthesis of GPI-anchored proteins. *J Am Chem Soc* **132**, 1567-1571 (2010).

## Chapter 1: Introduction

63. Wu, Z., Guo, X. & Guo, Z. Chemoenzymatic synthesis of glycosylphosphatidylinositol-anchored glycopeptides. *Chem Commun (Camb)* **46**, 5773-5774 (2010).
64. Antos, J.M., Miller, G.M., Grotenbreg, G.M. & Ploegh, H.L. Lipid modification of proteins through sortase-catalyzed transpeptidation. *J Am Chem Soc* **130**, 16338-16343 (2008).
65. Kobashigawa, Y., Kumeta, H., Ogura, K. & Inagaki, F. Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. *J Biomol NMR* **43**, 145-150 (2009).
66. Sakamoto, T., Sawamoto, S., Tanaka, T., Fukuda, H. & Kondo, A. Enzyme-Mediated Site-Specific Antibody-Protein Modification Using a ZZ Domain as a Linker. *Bioconjug Chem* (2010).
67. Matsushita, T. & Nishimura, S. Novel synthesis of functional mucin glycopeptides containing both N- and O-glycans. *Methods Enzymol* **478**, 485-502 (2010).
68. Matsushita, T., *et al.* Functional neoglycopeptides: synthesis and characterization of a new class of MUC1 glycoprotein models having core 2-based O-glycan and complex-type N-glycan chains. *Biochemistry* **48**, 11117-11133 (2009).
69. Pritz, S., *et al.* Synthesis of protein mimics with nonlinear backbone topology by a combined recombinant, enzymatic, and chemical synthesis strategy. *Angew Chem Int Ed Engl* **47**, 3642-3645 (2008).
70. Antos, J.M., *et al.* A straight path to circular proteins. *J Biol Chem* **284**, 16028-16036 (2009).
71. Clark, R.J., *et al.* Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. *Proc Natl Acad Sci U S A* **102**, 13767-13772 (2005).
72. Clark, R.J., *et al.* The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew Chem Int Ed Engl* **49**, 6545-6548 (2010).
73. Trabi, M. & Craik, D.J. Circular proteins--no end in sight. *Trends Biochem Sci* **27**, 132-138 (2002).
74. Craik, D.J. Chemistry. Seamless proteins tie up their loose ends. *Science* **311**, 1563-1564 (2006).
75. Andersen, A.S., *et al.* Backbone cyclic insulin. *J Pept Sci* **16**, 473-479 (2010).
76. Wu, S. & Proft, T. The use of sortase-mediated ligation for the immobilisation of bacterial adhesins onto fluorescence-labelled microspheres: a novel approach to analyse bacterial adhesion to host cells. *Biotechnol Lett* **32**, 1713-1718 (2010).
77. Chan, L., *et al.* Covalent attachment of proteins to solid supports and surfaces via Sortase-mediated ligation. *PLoS One* **2**, e1164 (2007).
78. Clow, F., Fraser, J.D. & Proft, T. Immobilization of proteins to biacore sensor chips using *Staphylococcus aureus* sortase A. *Biotechnol Lett* **30**, 1603-1607 (2008).
79. Ito, T., *et al.* Highly oriented recombinant glycosyltransferases: site-specific immobilization of unstable membrane proteins by using *Staphylococcus aureus* sortase A. *Biochemistry* **49**, 2604-2614 (2010).
80. Mao, H. A self-cleavable sortase fusion for one-step purification of free recombinant proteins. *Protein Expr Purif* **37**, 253-263 (2004).

Chapter 1: Introduction

81. Matsunaga, S., Matsuoka, K., Shimizu, K., Endo, Y. & Sawasaki, T. Biotinylated-sortase self-cleavage purification (BISOP) method for cell-free produced proteins. *BMC Biotechnol* **10**, 42 (2010).
82. Caswell, J., Snoddy, P., McMeel, D., Buick, R.J. & Scott, C.J. Production of recombinant proteins in *Escherichia coli* using an N-terminal tag derived from sortase. *Protein Expr Purif* **70**, 143-150 (2010).
83. Bentley, M.L., Gaweska, H., Kielec, J.M. & McCafferty, D.G. Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J Biol Chem* **282**, 6571-6581 (2007).



**Figure Legends**

**Figure 1.1. Site-specific C- and N-terminal labeling scheme using sortase A.**

C-terminal labeling (left) and N-terminal labeling (right) proceed through a substrate recognition step (top), followed by generation of a thioacyl intermediate (middle) and resolution of the acylated enzyme by an exogenously added nucleophile (bottom). See text for details.

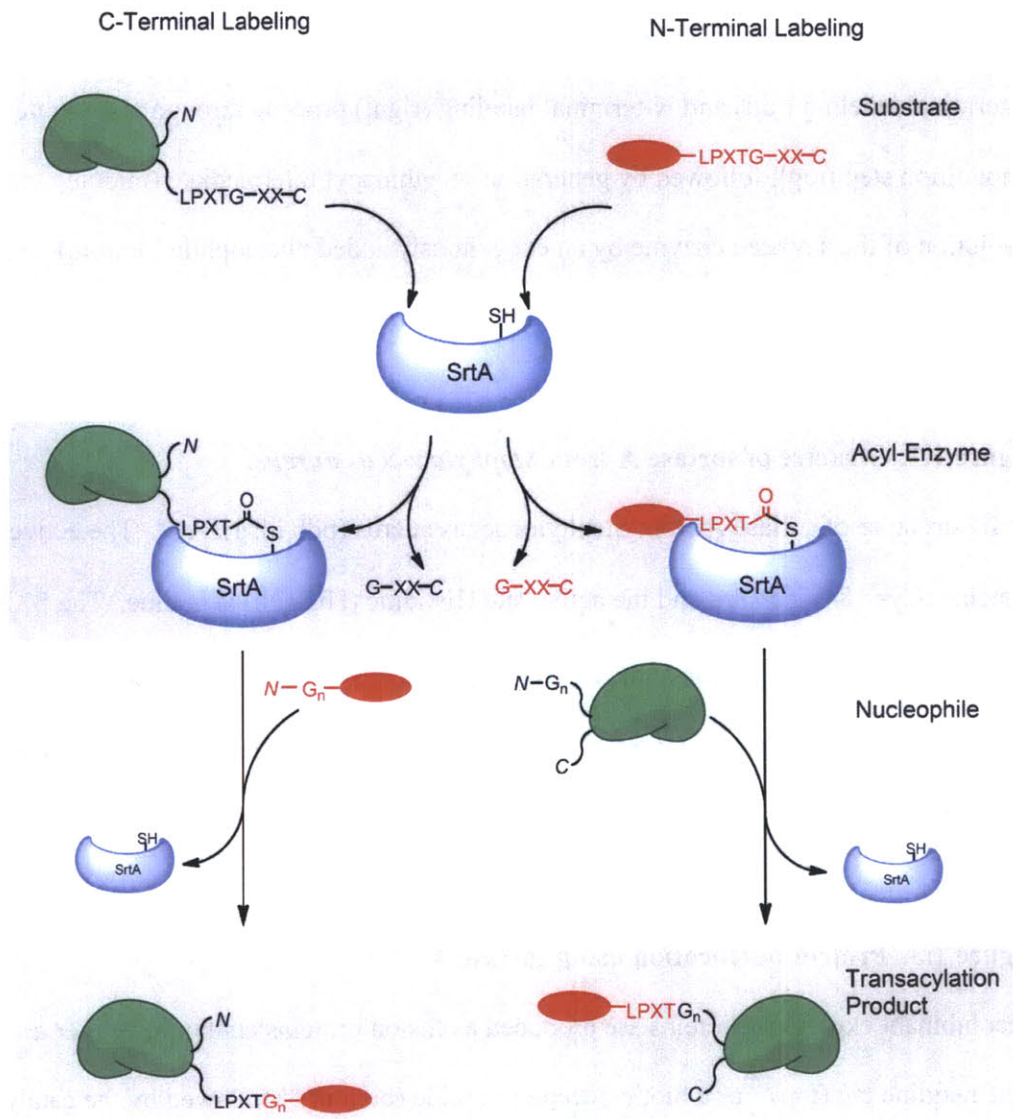
**Figure 1.2. Structure of sortase A from *Staphylococcus aureus*.**

NMR structure of sortase A from *Staphylococcus aureus* (pdb id: 1IJA)<sup>33</sup>. The active site cysteine (Cys 184) is in red and the active site Histidine (His 120) is in blue. The  $\beta$ 7 and  $\beta$ 8 strands that form the floor of the active site are labeled and the  $\beta$ 6- $\beta$ 7 loop, involved in substrate recognition<sup>83</sup> is in purple. Residues that coordinate calcium are shown as sticks<sup>39</sup>.

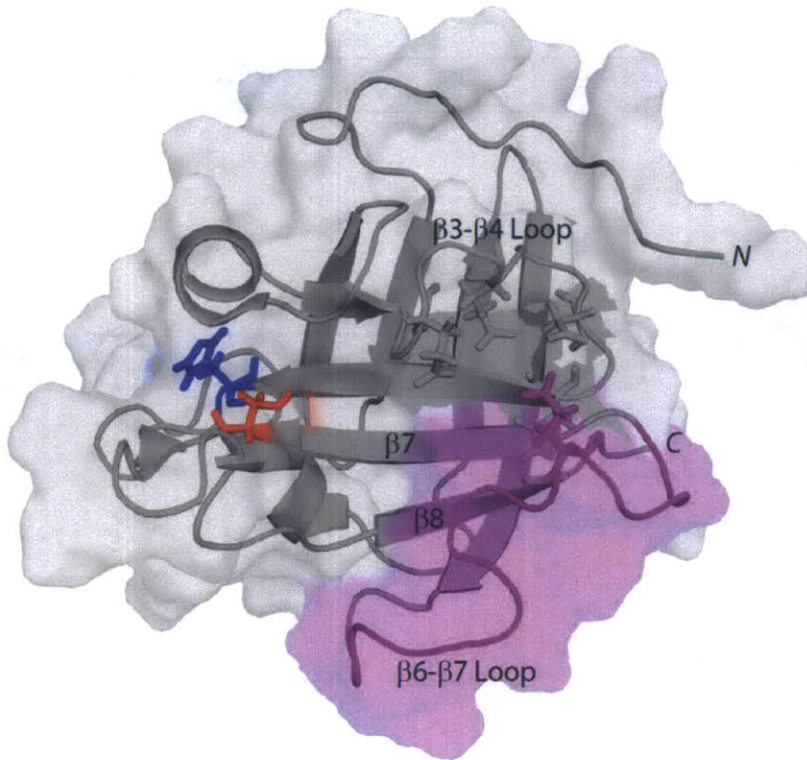
**Figure 1.3. Protein purification using sortase A.**

Recombinant expressed proteins are produced as fusion proteins containing either a hexahistidine tag (top)<sup>80</sup> or a biotin acceptor peptide (bottom)<sup>81</sup> followed by the catalytic core of sortase, the LPXTG tag, and the protein of interest. Addition of  $\text{Ca}^{2+}$  ions and oligoglycine to the immobilized fusion protein stimulates sortase activity. The protein of interest is released as a purified preparation with one additional N-terminal glycine.

Figure 1.1



**Figure 1.2**



**Figure 1.3.**

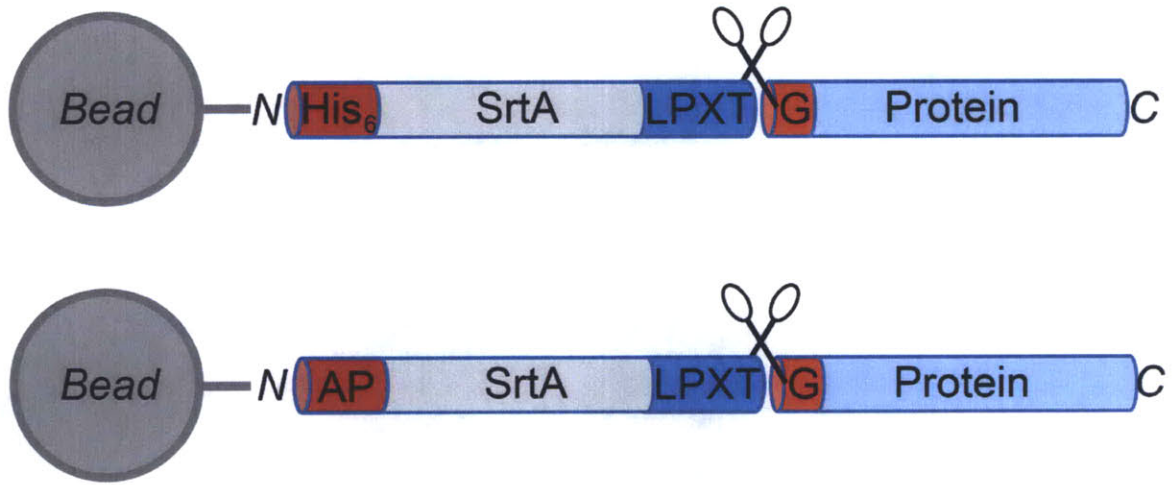
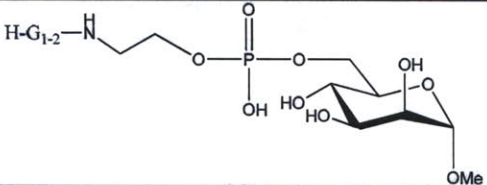


Table 1.1. Examples of synthetic nucleophiles used in site-specific sortase A transpeptidation reactions.

	Probe	N-or C-terminal Labeling	Property Endowed	Reference(s)
1	H-G <sub>5</sub> K(Biotin)L-OH	C-terminal	Biophysical handle	54, 53
2	H-G <sub>5</sub> K(ANP)K(Biotin)L-OH	C-terminal	Biophysical handle/Photocleavage	54
3	H-G <sub>5</sub> K(Phenylazide)K(Biotin)G-OH	C-terminal	Biophysical handle/Photocrosslinker	54
4	H-G <sub>3</sub> K(FITC)-NH <sub>2</sub>	C-terminal	Fluorescence	54
5	H-G <sub>3</sub> K(K(TAMRA))-NH <sub>2</sub>	C-terminal	Fluorescence	54, 70
6	H-G <sub>3</sub> YC(Biotin)-NH <sub>2</sub>	C-terminal	Biophysical Handle	55
7	H-G <sub>3</sub> YC(Alexa 488)-NH <sub>2</sub>	C-terminal	Fluorescence	55
8	H-AA-Ahx-K(K(TAMRA))-NH <sub>2</sub>	C-terminal ( <i>S. pyogenes</i> )	Fluorescence	56
9	H-G <sub>3</sub> K(C12-C24)-NH <sub>2</sub>	C-terminal	Lipidation	64
10	H-G <sub>3</sub> K(1-ad)-NH <sub>2</sub>	C-terminal	Hydrophobicity	64
11	H-G <sub>3</sub> WK(Cholesterol)-NH <sub>2</sub>	C-terminal	Lipidation	64
12	D-Tat (1 <sup>st</sup> residue is G)	C-terminal	Cell Penetration	51
13	H-G <sub>2</sub> Y-PTD5-NH <sub>2</sub>	C-terminal	Cell Penetration	51
14	(H <sub>2</sub> NRRQRRTSKLMKRAhx) <sub>2</sub> KYK(GG-NH <sub>2</sub> )-NH <sub>2</sub>	C-terminal	Cell Penetration	51
15	H-G <sub>3</sub> K(Folate)-NH <sub>2</sub>	C-terminal	Folic Acid	51
16	H <sub>2</sub> N-PEG	C-terminal	Inert Polymer	52
17	H-G <sub>3</sub> K(PEG)-OH	C-terminal	Inert Polymer	52
18	H-G <sub>3</sub> -MAP-NH <sub>2</sub>	C-terminal	Cell Penetration	50
19	Aminoglycoside antibiotics (various)	C-terminal	Antibiotic	60
20		C-terminal	GPI Mimic	61
21	GPI mimics based on 19 with trisaccharide cores	C-terminal	GPI Mimic	62, 63
22	Biotin-PEG-YGLPETGG-NH <sub>2</sub>	N-terminal	Biophysical handle	57
23	Alexa 647-LPETGG-NH <sub>2</sub>	N-terminal	Fluorescence	57
24	Alexa 488-LPETGG-NH <sub>2</sub>	N-terminal	Fluorescence	58
25	Biotin-LPRT-OMe	N-terminal	Biophysical handle	56
26	FITC-Ahx-LPRT-OMe	N-terminal	Fluorescence	56
27	FAM-LPETG-NH <sub>2</sub>	N-terminal	Fluorescence	47
28	Biotin-GGLPETG-NH <sub>2</sub>	N-terminal	Biophysical handle	47
29	N <sub>3</sub> -ALPETG-NH <sub>2</sub>	N-terminal	Handle for bioorthogonal chemical reactions	47

**Table 1.2. Examples of proteins labeled by sortase A transpeptidation.**

<b>Substrate</b>	<b>Solution/Cell Surface</b>	<b>N-or C-terminal Labeling</b>	<b>Label(s)</b>	<b>Reference(s)</b>
H-2K <sup>b</sup>	Solution	C-terminal	<b>1,2,3,4,5</b>	<sup>54</sup>
CD154	Cell Surface	C-terminal	<b>1,5</b>	<sup>54</sup>
Neuramidase	Cell Surface	C-terminal	<b>1</b>	<sup>54</sup>
ODF	Cell Surface	C-terminal	<b>6,7</b>	<sup>55</sup>
Cre	Solution	C-terminal	<b>5</b>	<sup>70</sup>
UCHL3	Solution	C-terminal (loop)	<b>1,5</b>	<sup>53</sup>
p97	Solution	C-terminal	<b>5</b>	<sup>70</sup>
eGFP	Solution	C-terminal	<b>9,10,11</b>	<sup>64</sup>
GFP	Solution	C-terminal	<b>13,14,15</b>	<sup>51</sup>
PNA	Solution	C-terminal	<b>18</b>	<sup>50</sup>
eGFP	Solution	C-terminal	<b>16,17</b>	<sup>52</sup>
Mrp	Solution	C-terminal	<b>19</b>	<sup>60</sup>
YALPETGK	Solution	C-terminal	<b>19</b>	<sup>60</sup>
(His) <sub>6</sub> YALPETGKS	Solution	C-terminal	<b>20</b>	<sup>61</sup>
CD52 Peptides	Solution	C-terminal	<b>21</b>	<sup>62</sup>
CD24	Solution	C-terminal	<b>21</b>	<sup>62</sup>
MUC1	Solution	C-terminal	<b>21</b>	<sup>63</sup>
LPETG <sub>5</sub> -ECFP-TM	Cell Surface	N-terminal	<b>22,23</b>	<sup>57</sup>
LPETG <sub>5</sub> -PAFR	Cell Surface	N-terminal	<b>24</b>	<sup>58</sup>
G <sub>3</sub> -/G <sub>5</sub> -CTXB	Solution	N-terminal	<b>25,26</b>	<sup>56</sup>
G <sub>3</sub> -eGFP	Solution	N-terminal	<b>26</b>	<sup>56</sup>
G-UCHL3	Solution	N-terminal	<b>26</b>	<sup>56</sup>
<i>S. aureus</i> Surface Peptidoglycan	Cell Surface	N-terminal	<b>27,28,29</b>	<sup>47</sup>
eGFP	Solution	N-and C-terminal	<b>26&amp;8</b>	<sup>56</sup>
UCHL3	Solution	N-and C-terminal	<b>26&amp;8</b>	<sup>56</sup>

## **Part I: Development of Sortase-mediated transpeptidation technology**

**Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells**



Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

**Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells**

This chapter is an expanded version of: Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**, 707-708 (2007).

**Abstract**

Genetically encoded reporter constructs that yield fluorescently labeled fusion proteins are a powerful tool for observing cell biological phenomena, but have their limitations. We describe a versatile chemo-enzymatic system for site-specific labeling of proteins, called sortagging (sortase-mediated transpeptidation), which combines the precision of a genetically encoded tag with the specificity of an enzymatic reaction and the ease and chemical versatility of peptide synthesis. We use sortase-mediated transpeptidation to install on proteins of interest a variety of small probes, readily accessible by standard peptide synthesis, which then allow interrogation by affinity-based methods or direct visualization by fluorescence. Sortagging may be applied in solution and on the surface of living cells, where the creation of GFP fusion proteins interferes with the function of the tagged protein and thus mandates the use of fluorophores of smaller size.

**Main Text**

The use of green fluorescent protein (GFP) and its derivatives has revolutionized the study of protein behavior in living cells <sup>1</sup>. It has further opened up the possibility of tracking cells bearing GFP-tagged proteins in living organisms. However, not all proteins tolerate the installation of GFP without compromising function or intracellular distribution <sup>2</sup>. Chemical methods for the installation of fluorescent or affinity labels have

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

the advantage of ease of use, but lack the precision of genetically encoded tags. To overcome this challenge, chemoenzymatic methods and small molecule binding peptide sequences that allow site-specific incorporation of labels have been developed, including transglutaminase-catalyzed reactions<sup>3</sup>, acyl carrier-protein-based labeling<sup>4</sup>, O6-alkylguanine DNA alkyltransferase fusions<sup>5</sup>, dihydrofolate reductase fusions<sup>6</sup>, biotin ligase<sup>7</sup>, FIAsh<sup>8</sup>, and Texas-red binding peptide<sup>9</sup>. However, several of these methods require the installation of a protein-sized module to afford selective labeling, or necessitate the insertion of recognition sequences that vary in size from 6 to 38 residues, with varying degrees of labeling selectivity. The nature of the labeling method also dictates the types of reporter molecules that can be installed, some of which require synthetic capabilities beyond the reach of most laboratories involved in biochemical or cell-biological studies. For example, the system that has perhaps enjoyed the widest application, FIAsh, is limited to installation of fluorophores, requires lengthy washouts to remove the biarsenical compound from endogenous monothiols<sup>10</sup>, and is not easily applied to proteins that traverse the secretory pathway unless free thiols are present or generated by reduction. Despite these limitations, the range of applications<sup>10-12</sup> demonstrates the utility of site-specific chemical labeling. Key features of these different labeling methods have been reviewed<sup>13</sup>.

Here we exploit the transpeptidase activity of bacterial sortases, thiol-containing enzymes that covalently attach proteins to the bacterial cell wall<sup>14</sup>, for selective labeling of proteins in solution, in cell lysates of complex composition, and on the surface of living cells, with a diverse set of probes suitable for the study of protein interactions and protein

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

trafficking. *Staphylococcus aureus* sortase A recognizes a set of structurally and functionally diverse substrates via an LPXTG motif, cleaving the peptide bond between threonine and glycine and subsequently forming an amide linkage with the N-terminus of a pentaglycine nucleophile provided *in vivo* by a cell wall precursor<sup>15</sup>. Sortase A has been used previously *in vitro* to affix cell permeable peptides<sup>16</sup> as well as poly(ethylene glycol) and polystyrene beads<sup>17</sup>, but the reactions were limited to a single model substrate, GFP. We use a set of oligoglycine-based nucleophiles and reaction conditions that allow efficient and selective labeling of biologically relevant protein complexes with affinity probes, photocrosslinkers, or fluorophores, both *in vitro* and on the surface of living cells (**Fig. 2.1**).

Expression of recombinant (His)<sub>6</sub> tagged sortase A in *E. coli* allows recovery in good yield (60 mg/liter of culture) of a highly purified preparation<sup>14</sup>. As a substrate, we generated a construct composed of the coding sequence for the luminal portion of the murine Class I MHC heavy chain H-2K<sup>b</sup>, followed by an LPETG motif and a C-terminal 15 amino acid acceptor peptide (AP) sequence for *E. coli* BirA biotin ligase. We expressed the recombinant H-2K<sup>b</sup> protein and reconstituted it with  $\beta_2$ -microglobulin ( $\beta_2m$ ) and the SIINFEKL peptide. We then biotinylated the AP portion of our refolded H-2K<sup>b</sup> using the BirA enzyme<sup>18</sup>. For sortagging, we incubated our LPETG-containing H-2K<sup>b</sup> monomers (15  $\mu$ M) bearing the biotinylated C-terminal AP tag with sortase A (150  $\mu$ M) in the presence or absence of biotinylated pentaglycine probe **1** (5 mM, **Fig. 2.2a**). In contrast to previous reports on the use of sortase-mediated transpeptidation to append oligoglycine derivatives to GFP<sup>16-17</sup>, excess sortase A was employed to drive the

transpeptidation reaction to completion within a reasonable period of time. In the absence of probe **1** we observed loss of biotin, corresponding to departure of the biotinylated AP tag and the concomitant formation of both a higher molecular weight and a lower molecular weight species (**Fig. 2.2a**). We excised the high molecular weight polypeptide from a silver stained gel and sequenced it after trypsinolysis by electrospray ionization tandem mass spectrometry (**Supplementary Figs. 2.1a and 2.1b**).

Remarkably, it was comprised of both sortase A and the H-2K<sup>b</sup> input substrate, indicating that it was in fact the acyl-enzyme intermediate<sup>19</sup>. Hydrolysis of this thioester intermediate generates the lower molecular weight band also observed in the absence of probe **1**. Upon addition of probe **1** to the reaction mixture, we observed rapid formation of the expected transpeptidation product: a sortase-cleaved, biotinylated H-2K<sup>b</sup> molecule lacking the 15 amino acid AP tag (**Fig. 2.2a**). The distinct electrophoretic mobility of the input substrate and the sortagged product allowed for convenient monitoring of the reaction by immunoblot with streptavidin–horseradish peroxidase (HRP). Quantitative replacement of the biotinylated AP tag with probe **1** occurred within 60 minutes.

To demonstrate that the reaction product was conformationally intact H-2K<sup>b</sup>, we tetramerized the quantitatively sortagged H-2K<sup>b</sup> by incubation with streptavidin-phycoerythrin (PE). Fully assembled SIINFEKL-loaded H-2K<sup>b</sup> tetramers will stain OT-1 cells, T cells from an ovalbumin specific H-2K<sup>b</sup> restricted T cell receptor transgenic line of mice<sup>20</sup>. Our sortagged tetramers indeed retained functional integrity, as assessed by robust fluorescent staining of OT-1 TCR transgenic T cells (**Fig. 2.2b**). Therefore, neither the LPETG tag nor the transpeptidation reaction perturbs the fold of this protein

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

complex. Because very little is known about the structural requirements for recognition of the LPXTG motif in natural bacterial substrates, we investigated the requirements for placement of the LPETG tag within the defined structure of the H-2K<sup>b</sup> protein.

Additional H-2K<sup>b</sup> substrates were prepared in which the LPETG motif was moved from the C-terminus to seven surface-exposed loop regions (**Supplementary Fig. 2.3a** and **Supplementary Table 2.1**). Proper folding of these substrates was largely retained, as indicated by their ability to stain OT-1 T cells following tetramerization (**Supplementary Fig. 2.3b**). However, upon exposure to sortase A and probe **1**, no transpeptidation was detected (**Supplementary Fig. 2.3c**). Thus, we conclude that the LPETG tag must be placed in a flexible, unstructured region close to the C-terminus of the substrate protein.

With the reaction conditions for quantitative labeling in hand, we next synthesized additional oligoglycine nucleophiles (probes **2-5**) compatible with the sortagging system and demonstrated their ability to efficiently label the H-2K<sup>b</sup> substrate (**Fig. 2.3**). A probe in which the biotin moiety is separated from the (Gly)<sub>5</sub> motif by a 3-amino-3-(*o*-nitrophenyl) propionic acid residue (probe **2**) allows photocleavage of the tag, with concomitant release of the biotin label. Sortagging can thus be reversed photochemically under mild conditions (**Fig. 2.4a**). We also generated a probe containing both a biotin moiety and an aryl azide photocrosslinker appended to the (Gly)<sub>5</sub> motif (probe **3**) and found this to be sufficient to crosslink the sortagged H-2K<sup>b</sup> heavy chain to  $\beta_2$ -microglobulin (**Fig. 2.4b**). Finally, for fluorescent visualization, we successfully installed either a fluorescein (FITC) or tetramethyl rhodamine (TAMRA) dye by means of the correspondingly modified triglycine peptide (probes **4** and **5** respectively). All of

the probes described here can be quantitatively appended with similar efficiencies, were accessed through standard solid phase peptide synthesis with commercially available building blocks, and illustrate the diversity of functionalities that can be used in the sortagging system. In addition to our H-2K<sup>b</sup> substrate, we have also successfully labeled an LPETG tagged version of the chemokine CXCL14 in solution (**Supplementary Fig. 2.2**).

Having shown the feasibility of sortagging purified proteins in solution, we turned to labeling substrates in complex mixtures of proteins in order to demonstrate the specificity of the transpeptidation reaction. We chose as a model substrate the human CD154 (CD40L) molecule, a type II membrane glycoprotein present on activated T lymphocytes and platelets<sup>21</sup>, which functions as a ligand for CD40. CD154-induced CD40 signaling is essential for a productive interaction between T and B cells to bring about class switch recombination and somatic hypermutation<sup>22-23</sup> and also correlates with interactions between dendritic cells and T cells<sup>24</sup>. We installed at the C-terminus of CD40L an LPETG tag preceded by a short flexible linker and followed by the HA epitope tag, and expressed this construct in HEK 293T cells. The CD40L species was the only polypeptide labeled by sortagging *in vitro* on cell lysates, as detected by subsequent immunoblotting using streptavidin-HRP (**Supplementary Fig. 2.4a**). Similarly, we have also sortagged soluble H-2K<sup>b</sup> complexes when present in a crude bacterial lysate (**Supplementary Fig. 2.4b**).

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

Cell surface molecules play a key role in signal transduction and cell-cell communication. Thus, we sought to demonstrate that the sortagging method can label protein substrates on the surface of live cells for direct visualization. HEK 293T cells transfected with the CD40L construct were incubated with serum-containing media and sortase A together with probe **1**. Upon incubation, sortagged CD40L was readily visualized with streptavidin-HRP, with almost no labeling of endogenous polypeptides (**Fig. 2.5a, Supplementary Fig. 2.5**). We detected the persistence of HA-tagged CD40L, indicating that sortase A does not attack the entire pool of CD40L substrate, since transpeptidation is accompanied by loss of the C-terminal HA tag (**Fig. 2.1**). This could be because not all CD40L is expressed at the cell surface in HEK 293T transfectants, or because only a portion of all surface CD40L is accessible to sortase. We cotransfected HEK293T cells with tagged CD40L substrate and a construct that specifies soluble eGFP. We detected specific labeling with the TAMRA-containing probe **5** only on GFP<sup>+</sup> (transfected) cells, even after an exposure to the labeling mixture for a mere 10 minutes (**Fig. 2.5b**). In addition to the CD40L molecule, we have also successfully labeled the surface-displayed influenza A/WSN/33 neuraminidase, the enzyme responsible for release of new virions from host cells (**Supplementary Fig. 2.6**). These experiments demonstrate the feasibility of rapid and highly selective chemo-enzymatic labeling of proteins with a suitably exposed C-terminal LPETG tag on the surface of living cells. Our results show that few if any endogenous mammalian proteins are attacked by sortase A. For labeling to occur, target proteins must be genetically tagged, a trait that further enhances the utility of sortagging.

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

Sortagging provides a robust and general method for site-specific protein labeling and unlike other labeling methods, allows single-step incorporation of a variety of probes well-suited for the study of protein interactions and protein trafficking via a stable amide linkage. Recombinant (His)<sub>6</sub> tagged sortase A can be readily produced in good yield and purity and, because of the nature of the transpeptidation reaction mechanism, may be simultaneously removed from crude reaction mixtures along with unreacted substrate if a (His)<sub>6</sub> tag is included C-terminal to the requisite LPXTG motif, to afford a pure probe-bearing species. When used on intact cells, simple washing effectively removes sortase A. In addition, the oligoglycine probes themselves are accessible by standard solid phase methods for peptide synthesis. The sortagging method described here derives its utility from the exploitation of a genetically encoded tag of only 5 residues<sup>25-26</sup> preceded by a short spacer. The multiplicity of sortases and their distinct recognition sequences<sup>27</sup> in principle immediately expands the range of labeling possibilities to the simultaneous use of several uniquely tagged proteins and various probes in a single experimental setting.

## Methods

**H-2K<sup>b</sup> Labeling.** Soluble H-2K<sup>b</sup> monomers were first biotinylated on the AP region by incubation with BirA enzyme in 50 mM bicine buffer, pH 8.3, 10 mM ATP, 10 mM MgOAc, and 50 μM d-biotin at 4° C overnight. The complexes were purified by size exclusion chromatography on S75 Sephadex resin using 20mM Tris, 50mM NaCl, pH 8.0



Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

buffer as eluent. Purified, soluble H-2K<sup>b</sup> monomers (15 μM) were labeled by incubation with 150 μM SrtA, 5 mM of the indicated probe in sortase buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) at 37°C. At the indicated time points, aliquots were removed and the reaction was halted by addition of reducing, denaturing SDS-PAGE sample buffer. For labeling of H-2K<sup>b</sup> monomers with LPETG substituted in loop regions, 500 μM probe was used. For fluorescence labeling, gels were visualized by scanning with a Typhoon Imager (GE Healthcare).

**H-2K<sup>b</sup> Tetramerization and FACS analysis.** For H-2K<sup>b</sup> with the LPETG tag at the C-terminus followed by the BirA AP, monomers were labeled to completion under the same conditions as in **Fig. 2.2** with probe 1 at 15 mM. Labeled monomers were then subjected to gel filtration over Sephadex S-75 resin and concentrated. Sortagged monomers (30 μg) were tetramerized by 5 hourly additions of 9 μg each (45 μg total) of streptavidin-PE on ice. Sortagged tetramers, propidium iodide (PI), and anti-CD8 antibody conjugated to FITC (BD Biosciences) were used to stain splenocytes isolated from OT-1 and 2C mice. FACS analysis was performed with a BD LSR flow cytometer. Dead cells (PI high) were gated out and live cells were displayed. H-2K<sup>b</sup> loop constructs were biotinylated on the BirA AP as described, subjected to gel filtration (Sephadex S-75) and concentrated. Constructs (8 μg) were tetramerized by 5 hourly additions of 2.4 μg (12 μg total) of streptavidin-PE on ice. Tetramers, propidium iodide (PI), and anti-CD8 antibody conjugated to FITC (BD Biosciences) were used to stain splenocytes isolated from OT-1

and Balb/C mice. FACS analysis was performed with a BD LSR flow cytometer. Dead cells (PI high) were gated out and live cells were displayed.

**Photocleavage and Photocrosslinking of Sortagged H-2K<sup>b</sup> (probes 2 and 3).** A large scale labeling (144  $\mu$ L total reaction volume) of soluble H-2K<sup>b</sup> monomers with probes 2 and 3 was performed following the standard protocol described above. For these reactions, the BirA AP region of H-2K<sup>b</sup> was not biotinylated prior to sortagging. The ligation products were then purified by size exclusion chromatography (S75 Sephadex; 20 mM Tris, 50 mM NaCl, pH 8.0 buffer). The labeled H-2K<sup>b</sup> monomers were concentrated by centrifugal ultrafiltration to a final concentration of  $\sim$ 10  $\mu$ M in 20 mM Tris pH 8.0, 50 mM NaCl as determined by Bradford assay. For photocleavage of H-2K<sup>b</sup>-(Probe 2) the ligation product, purified H-2K<sup>b</sup>-(Probe 2) (1.0  $\mu$ L of a 10  $\mu$ M stock) and 8.0  $\mu$ L of 50 mM phosphate pH 6.5 containing 6.25 mM DTT as a scavenger for reactive by-products of photocleavage<sup>28</sup> were mixed. Biotinylated, uncleaved H-2K<sup>b</sup> (1.0  $\mu$ L of a 10  $\mu$ M stock in 20 mM Tris pH 8.0, 50 mM NaCl) was then added to provide an internal reference that was resistant to photocleavage. Five individual reactions were prepared, placed on ice, and irradiated for 0, 15, 30, 60 and 90 min (one reaction for each time point) in a UV Stratalinker 2400 crosslinker (Stratagene, USA) fitted with lamps emitting at 365 nm. Reactions were terminated with 40  $\mu$ L of reducing, denaturing SDS-PAGE sample buffer. Immunoblot analysis with streptavidin-HRP revealed no loss of biotin label for uncleaved H-2K<sup>b</sup> and significant loss of biotin label for H-2K<sup>b</sup>-(Probe 2) (**Figure 2.4a**). The maximum level of photocleavage was observed at the first time point (15 min), and no significant increase in the level of photocleavage was detected with

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

longer irradiation times. For photocrosslinking of H-2K<sup>b</sup>-(Probe 3) Ligation Product to  $\beta_2m$ , purified H-2K<sup>b</sup>-(Probe 3) (1.0 uL of a 10  $\mu$ M stock) and 9.0 uL of 50 mM phosphate pH 8.0 were mixed. Five individual reactions were prepared, placed on ice, and then irradiated for 0, 15, 30, 60 and 90 min (one reaction for each time point) in a UV Stratalinker 2400 crosslinker (Stratagene, USA) fitted with lamps emitting at 365 nm. Reactions were terminated with 40 uL of reducing, denaturing SDS-PAGE sample buffer. Immunoblot analysis with streptavidin-HRP revealed the appearance of a higher molecular weight species only after UV irradiation (**Figure 2.4b**). The apparent molecular weight increase from H-2K<sup>b</sup>-(Probe 3) was appropriate for crosslinking to  $\beta_2m$  (MW ~11 kD). No additional characterization of the crosslinked species was performed. The maximum level of crosslinking was observed at the first time point (15 min) and no significant increase in the level of crosslinking was detected with longer irradiation times.

**Cell Surface Labeling.** Approximately 12-24 hours after transfection, cells were labeled by incubating in DME/10 % IFS media containing 200  $\mu$ M sortase A and 500  $\mu$ M biotinylated probe 1 at 37° C. At the indicated time points, media was aspirated and cells were washed and harvested by scraping into ice-cold PBS. Labeled cells were pelleted and washed repeatedly with ice-cold PBS. Cells were lysed in 1 % SDS/PBS containing complete mini-protease inhibitor (Roche) and protein concentration was determined by BCA assay (Pierce). 20  $\mu$ g of lysate was loaded onto 12.5 % SDS-PAGE gel for immunoblot.

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

**Cell Imaging.** Approximately 12-24 hours after transfection, cells were trypsinized, replated onto poly-lysine coated 18 mm circular glass coverslips (VWR International) in a 12-well dish at a density of  $3 \times 10^5$  cells/well, and allowed to attach overnight. Cells were labeled by inverting coverslips onto a drop of 50  $\mu$ l of DME/10 % IFS media containing 200  $\mu$ M SrtA and 100  $\mu$ M probe **5** on parafilm. After 10 minutes of incubation at 37° C, coverslips were removed and washed repeatedly in ice-cold PBS with 1 mM MgCl<sub>2</sub> and 1 mM CaCl<sub>2</sub>. Cells were fixed by inversion onto a drop of 4 % paraformaldehyde in PBS for 10 minutes at room temperature on parafilm. Coverslips were removed and again washed repeatedly in ice-cold PBS with 1 mM MgCl<sub>2</sub> and 1 mM CaCl<sub>2</sub>. Coverslips were mounted with Fluoromount G (Southern Biotech) and analyzed using a Nikon spinning disk confocal microscope with Metamorph software. Multidimensional acquisition was used to collect images in the 488-, 568-, and DIC channels from the same focal plane.

**Acknowledgements**

We thank Dr. Olaf Schneewind for the generous gift of the sortase A encoding plasmid. We would like to thank Benjamin J. Hackel and Dr. K. Dane Wittrup for the generation of the tagged EGFR fragment and Dr. Ching-Hung Shen and Dr. Jianzhu Chen for the neuraminidase template. We also thank members of the Ploegh lab and Dr. Howard Hang for helpful discussions. This work was supported by grants from the National Institutes of Health.

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

**References**

1. Lippincott-Schwartz, J. & Patterson, G.H. Development and use of fluorescent protein markers in living cells. *Science* **300**, 87-91 (2003).
2. Lisenbee, C.S., Karnik, S.K. & Trelease, R.N. Overexpression and mislocalization of a tail-anchored GFP redefines the identity of peroxisomal ER. *Traffic* **4**, 491-501 (2003).
3. Lin, C.W. & Ting, A.Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J Am Chem Soc* **128**, 4542-4543 (2006).
4. George, N., Pick, H., Vogel, H., Johnsson, N. & Johnsson, K. Specific labeling of cell surface proteins with chemically diverse compounds. *J Am Chem Soc* **126**, 8896-8897 (2004).
5. Keppler, A., *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **21**, 86-89 (2003).
6. Miller, L.W., Sable, J., Goelet, P., Sheetz, M.P. & Cornish, V.W. Methotrexate conjugates: a molecular in vivo protein tag. *Angew Chem Int Ed Engl* **43**, 1672-1675 (2004).
7. Chen, I., Howarth, M., Lin, W. & Ting, A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2**, 99-104 (2005).
8. Griffin, B.A., Adams, S.R. & Tsien, R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272 (1998).
9. Marks, K.M., Rosinov, M. & Nolan, G.P. In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem Biol* **11**, 347-356 (2004).
10. Sun, M., *et al.* Dynamics of the upper 50-kDa domain of myosin V examined with fluorescence resonance energy transfer. *J Biol Chem* **281**, 5711-5717 (2006).
11. Gaietta, G., *et al.* Multicolor and electron microscopic imaging of connexin trafficking. *Science* **296**, 503-507 (2002).
12. Roberti, M.J., Bertoncini, C.W., Klement, R., Jares-Erijman, E.A. & Jovin, T.M. Fluorescence imaging of amyloid formation in living cells by a functional, tetracysteine-tagged alpha-synuclein. *Nat Methods* (2007).
13. Chen, I. & Ting, A.Y. Site-specific labeling of proteins with small molecules in live cells. *Curr Opin Biotechnol* **16**, 35-40 (2005).
14. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci US A* **96**, 12424-12429 (1999).
15. Marraffini, L.A. & Schneewind, O. Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Mol Microbiol* **62**, 1402-1417 (2006).
16. Mao, H., Hart, S.A., Schink, A. & Pollok, B.A. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **126**, 2670-2671 (2004).
17. Parthasarathy, R., Subramanian, S. & Boder, E.T. Sortase A as a Novel Molecular "Stapler" for Sequence-Specific Protein Conjugation. *Bioconjug Chem* (2007).

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

18. Beckett, D., Kovaleva, E. & Schatz, P.J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **8**, 921-929 (1999).
19. Aulabaugh, A., *et al.* Development of an HPLC assay for *Staphylococcus aureus* sortase: Evidence for the formation of the kinetically competent acyl enzyme intermediate. *Anal Biochem* **360**, 14-22 (2007).
20. Skinner, P.J., Daniels, M.A., Schmidt, C.S., Jameson, S.C. & Haase, A.T. Cutting edge: In situ tetramer staining of antigen-specific T cells in tissues. *J Immunol* **165**, 613-617 (2000).
21. Henn, V., *et al.* CD40 ligand on activated platelets triggers an inflammatory reaction of endothelial cells. *Nature* **391**, 591-594 (1998).
22. Grewal, I.S. & Flavell, R.A. CD40 and CD154 in cell-mediated immunity. *Annu Rev Immunol* **16**, 111-135 (1998).
23. Lougaris, V., Badolato, R., Ferrari, S. & Plebani, A. Hyper immunoglobulin M syndrome due to CD40 deficiency: clinical, molecular, and immunological features. *Immunol Rev* **203**, 48-66 (2005).
24. Hernandez, M.G., Shen, L. & Rock, K.L. CD40-CD40 Ligand Interaction between Dendritic Cells and CD8<sup>+</sup> T Cells Is Needed to Stimulate Maximal T Cell Responses in the Absence of CD4<sup>+</sup> T Cell Help. *J Immunol* **178**, 2844-2852 (2007).
25. Kruger, R.G., *et al.* Analysis of the substrate specificity of the *Staphylococcus aureus* sortase transpeptidase SrtA. *Biochemistry* **43**, 1541-1551 (2004).
26. Zong, Y., Bice, T.W., Ton-That, H., Schneewind, O. & Narayana, S.V. Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J Biol Chem* **279**, 31383-31389 (2004).
27. Dramsi, S., Trieu-Cuot, P. & Bierne, H. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. *Res Microbiol* **156**, 289-297 (2005).
28. Rinnova, M., Novakova, M., Kasicka, V. & Jiracek, J. Side reactions during photochemical cleavage of an alpha-methyl-6-nitroveratryl-based photolabile linker. *J Pept Sci* **6**, 355-365 (2000).
29. Garboczi, D.N., Hung, D.T. & Wiley, D.C. HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in *Escherichia coli* and complexed with single antigenic peptides. *Proc Natl Acad Sci U S A* **89**, 3429-3433 (1992).
30. Sleeman, M.A., *et al.* B cell- and monocyte-activating chemokine (BMAC), a novel non-ELR {alpha}-chemokine 10.1093/intimm/12.5.677. *Int. Immunol.* **12**, 677-689 (2000).

## Figure Legends

### Figure 2.1. Scheme for sortagging of proteins in solution and on the cell surface.

Proteins equipped with the LPETG tag are cleaved by sortase A at the threonine-glycine peptide bond. The glycine and residues C-terminal of it are released from the substrate protein with concomitant formation of a thioester-linked acyl-enzyme intermediate. This intermediate is resolved by a suitably functionalized oligoglycine nucleophile, resulting in the selective labeling of the protein substrate.

### Figure 2.2. Sortase A-mediated transpeptidation quantitatively labels the H-2K<sup>b</sup> complex *in vitro*.

(a) H-2K<sup>b</sup> monomers (15  $\mu$ M) with a C-terminal LPETG tag followed by a fully biotinylated 15 amino acid BirA acceptor peptide tag were incubated with 150  $\mu$ M sortase A in the presence or absence of 5 mM probe 1 for 1 hour at 37° C and resolved by SDS-PAGE for both coomassie staining and streptavidin-HRP immunoblot. In the absence of probe 1, the acyl-enzyme intermediate is formed, along with the H-2K<sup>b</sup> hydrolysis product indicated (lane 4). In the presence of sortase A and probe 1, a species of lower molecular weight (relative to input material) corresponding to the transpeptidation product (which lacks the 15 amino acid C-terminal BirA AP tag) is formed (lanes 5-7). Both the input H-2K<sup>b</sup> substrate that is biotinylated on the BirA AP tag and the lower molecular weight transpeptidation product containing biotinylated probe 1 are revealed by streptavidin-HRP immunoblot (compare lanes 2-4 and 5-7).

(b) FACS analysis of tetramerized sortagged H-2K<sup>b</sup> molecules labeled to completion with probe 1 and loaded with SIINFEKL peptide. Sortagged tetramers were used to stain

both OT-1 splenocytes (left) and 2C splenocytes (right), which recognize a distinct peptide-MHC combination<sup>20</sup>.

**Figure 2.3. Sortase-mediated transpeptidation allows site-specific incorporation of an array of probes suitable for biophysical studies and for direct visualization.**

Biotinylated pentaglycine probes bearing a photocleavable 3-amino-3-(*o*-nitrophenyl) propionic acid residue (**2**) or an aryl azide for photocrosslinking (**3**) are quantitatively appended to H-2K<sup>b</sup> under the same conditions as in **Fig. 1**. Fluorescent triglycine derivatives containing FITC (**4**) or TAMRA (**5**) allow direct visualization of transpeptidation products.

**Figure 2.4. H-2K<sup>b</sup> sortagged with probes 2 and 3 were found to undergo photocleavage or crosslinking, respectively, upon irradiation with UV light.**

(a) A mixture of sortagged H-2K<sup>b</sup>-(Probe **2**) and uncleaved, biotinylated H-2K<sup>b</sup> input material was subjected to UV irradiation and analyzed by streptavidin immunoblot. Loss of the biotin label was observed only for the sortagged conjugate containing the photocleavable 3-amino-3-(*o*-nitrophenyl) propionic acid functional group.

(b) UV irradiation of sortagged H-2K<sup>b</sup>-(Probe **3**) resulted in the formation of a higher molecular weight band consistent with covalent crosslinking of  $\beta_2$ -microglobulin to the H-2K<sup>b</sup> heavy chain.

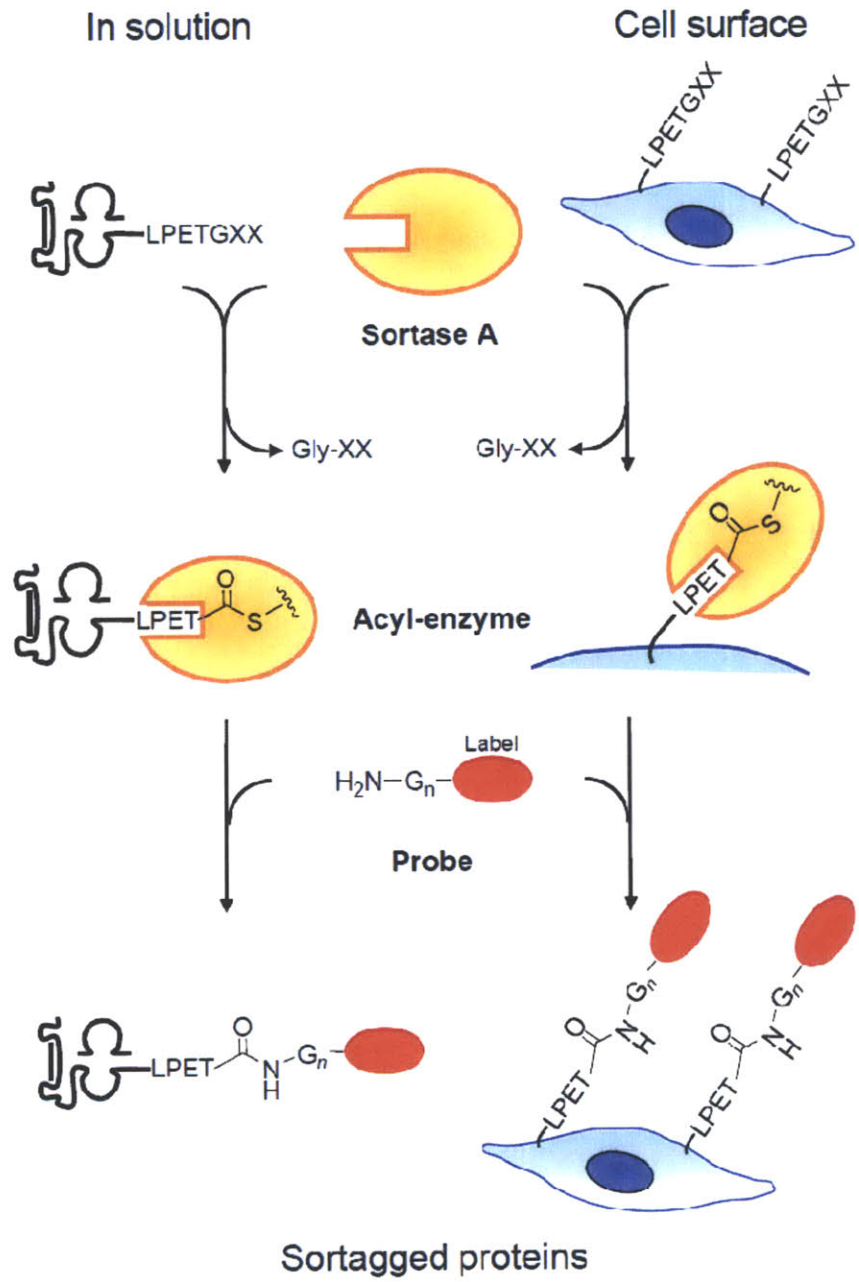


**Figure 2.5. Sortagging LPETG-bearing proteins on the surface of live cells.**

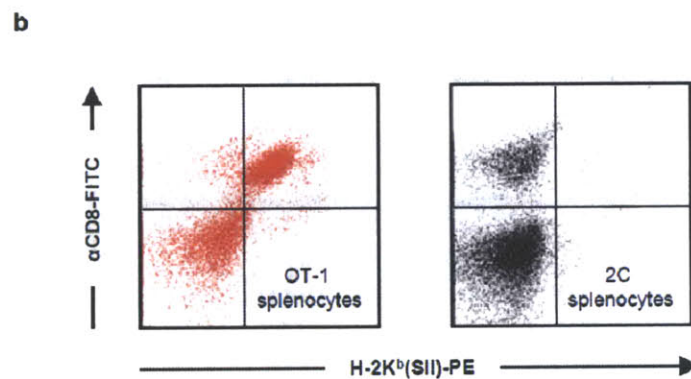
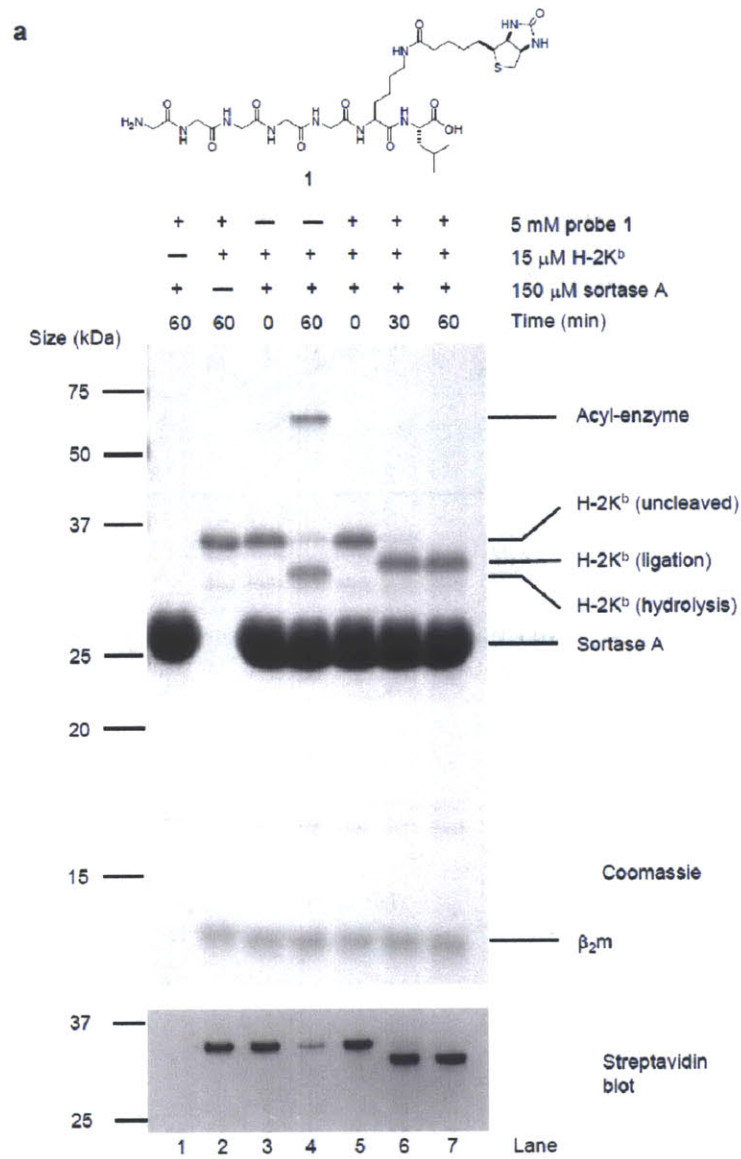
(a) CD154 (CD40L) was fused to a flexible linker followed by the LPETG tag and an HA epitope at the C-terminus and transfected into HEK 293T cells. Transfected and untransfected cells were incubated in serum-containing media with sortase A and probe **1** for the indicated times and collected for analysis by streptavidin-HRP and anti-HA immunoblot.

(b) Cells cotransfected with plasmids encoding soluble, cytoplasmic eGFP and the sortagged CD154 molecule were plated on glass coverslips, incubated with sortase A and probe **5** for 10 minutes, fixed, and imaged by spinning-disk confocal microscopy. (Scale bars, 10  $\mu$ M).

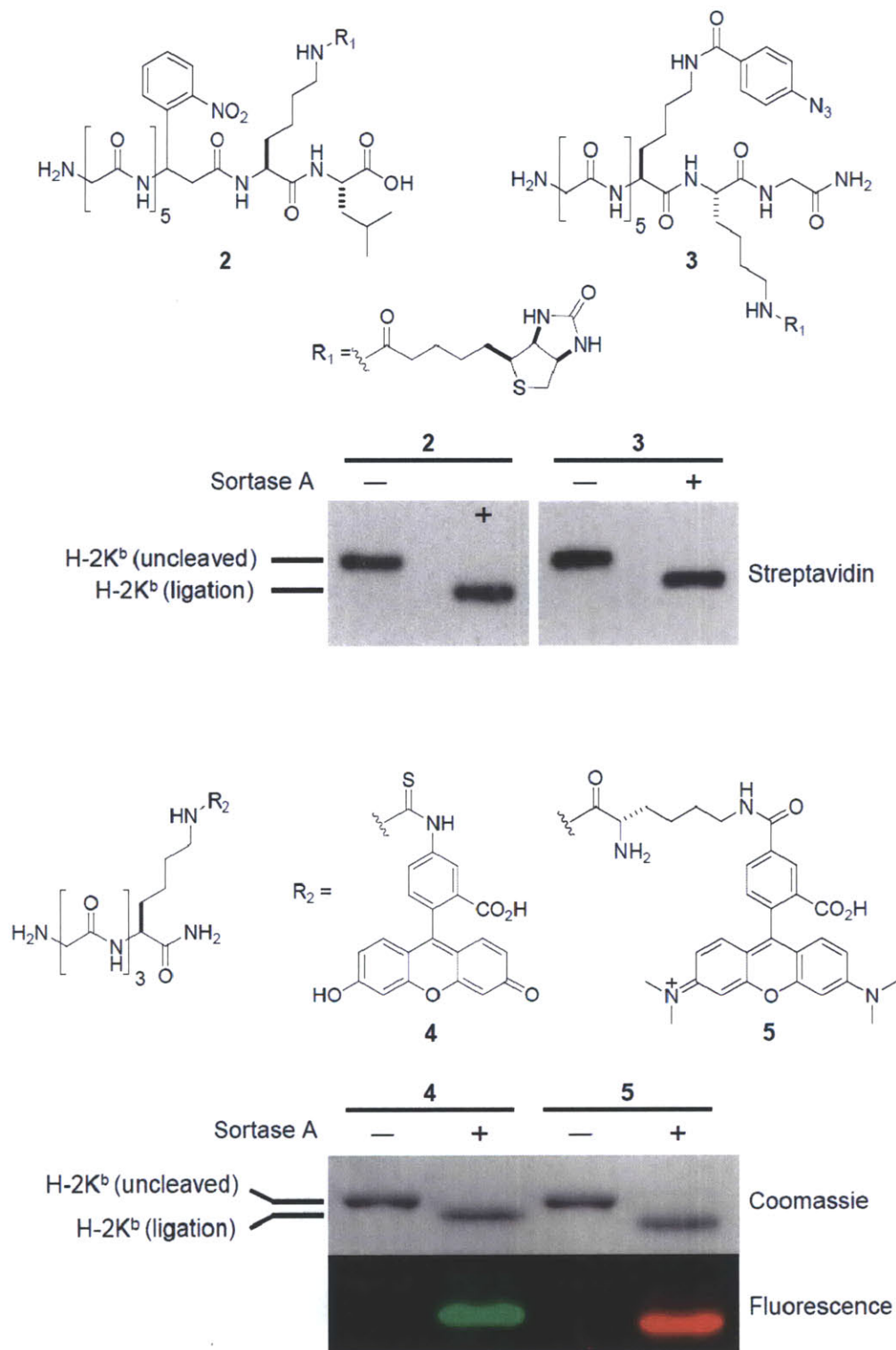
**Figure 2.1**



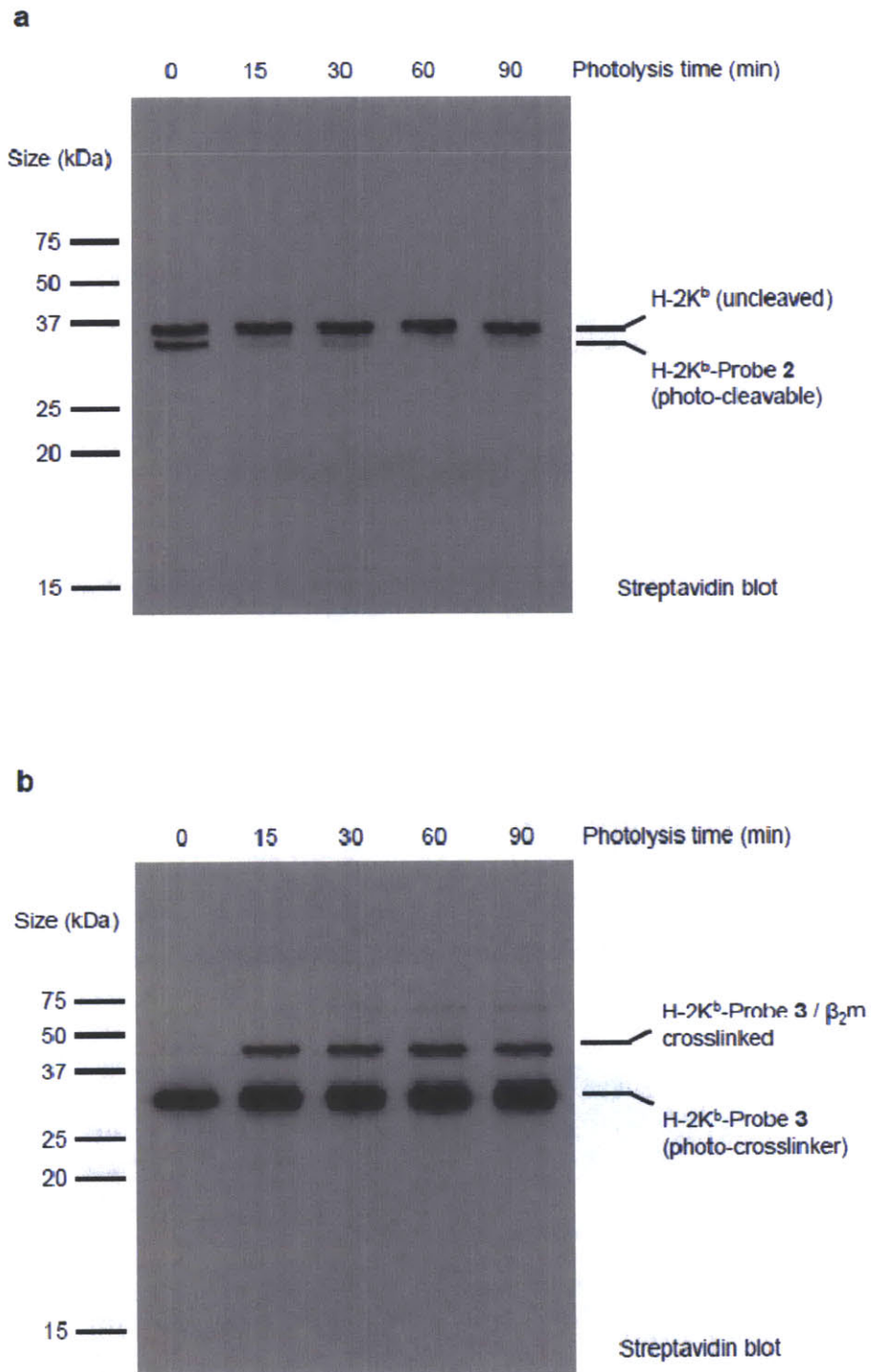
**Figure 2.2**



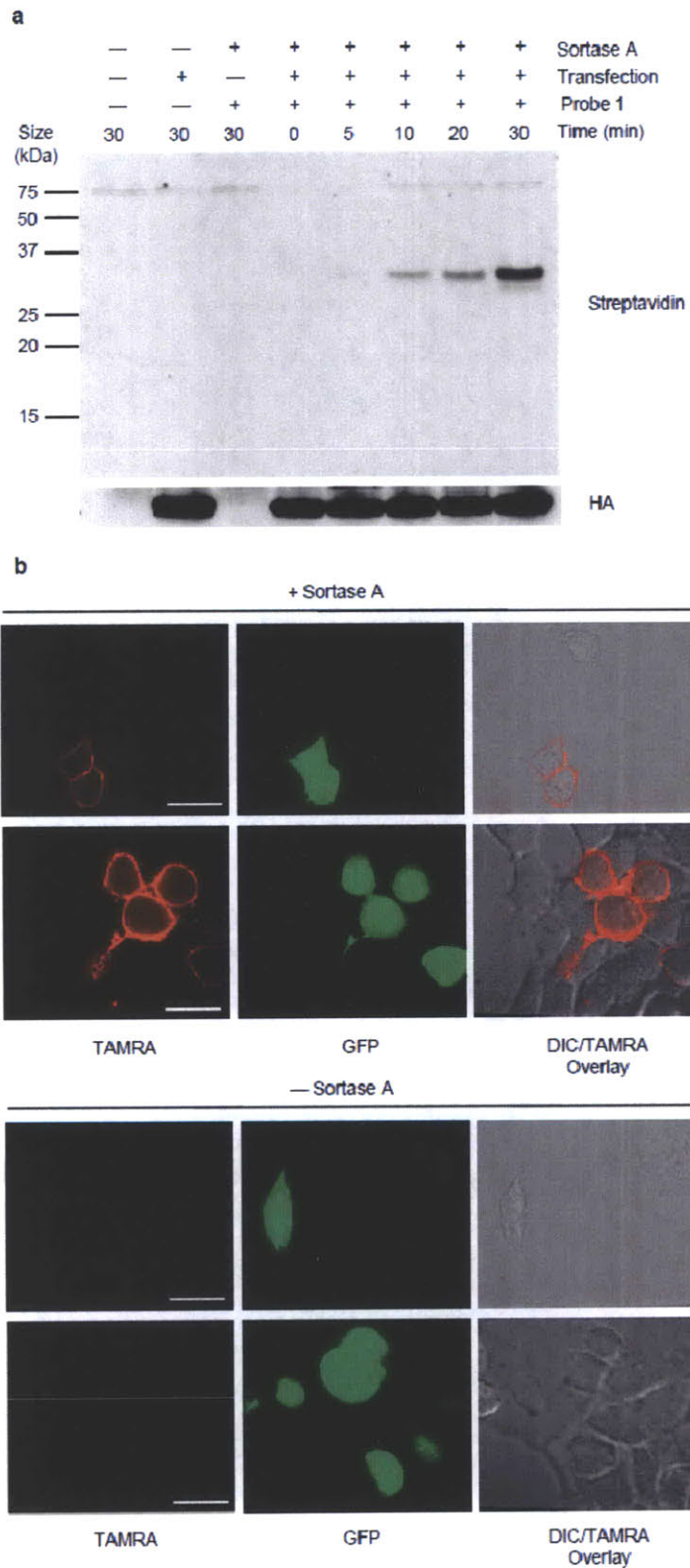
**Figure 2.3**



**Figure 2.4**



**Figure 2.5**



## Supplementary Figure Legends

**Supplementary Figure 2.1 The high molecular weight adduct formed during incubation of H-2K<sup>b</sup> with sortase A in the absence of oligoglycine probe is the acyl-enzyme intermediate.**

Excision of the high molecular weight band from a silver stained gel followed by trypsinolysis and electrospray ionization tandem mass spectrometry of the resulting peptides allow the positive identification of a number of peptides derived from (a) *S. aureus* sortase A and (b) H-2K<sup>b</sup>

**Supplementary Figure 2.2 Sortagging LPETG-bearing human CXCL14.**

LPETG-fused CXCL14 was treated with 150 μM sortase A and 5 mM **5** for 22 hours at RT and analyzed by SDS-PAGE.

**Supplementary Figure 2.3 Substitution of the LPETG tag into each of 7 defined loop regions in H-2K<sup>b</sup> renders these molecules refractory to cleavage by sortase A.**

(a) The LPETG tag was substituted into each of the loop regions shown in red and each H-2K<sup>b</sup> construct was fused at the C-terminus to the BirA AP (**Supplementary Table 1**).

(b) H-2K<sup>b</sup> loop constructs were biotinylated on the BirA AP, tetramerized by incubation with streptavidin-PE, and used to stain splenocytes from OT-1 and Balb/C (H-2<sup>d</sup>) mice.

The loop 7 replacement was non-functional. (c) Reconstituted H-2K<sup>b</sup> loop constructs were incubated in the presence of sortase A and probe **1** for 1 hour at 37° C. As a positive control, the refolded H-2K<sup>b</sup> monomer bearing a C-terminal LPETG tag followed by a 15

amino acid BirA AP tag is also included. Reaction mixtures were resolved by SDS-PAGE and stained with coomassie.

**Supplementary Figure 2.4 Sortagging installs probes onto suitably tagged molecules in crude mammalian and bacterial lysates.**

(a) HEK 293T cells were transfected with a plasmid expressing the CD154 molecule C-terminally fused to a flexible linker followed by the LPETG tag and an HA epitope.

Transfected and untransfected cells were lysed in NP-40 (0.5%) and lysates were incubated in the presence of probe **1** with or without sortase A for 1 hour at 37° C.

Reactions were resolved by SDS-PAGE for both streptavidin-HRP and anti-HA immunoblot. Reactions were also resolved by SDS-PAGE and gels were stained with coomassie to reveal the complexity of polypeptides in the lysates.

(b) BL-21 *E. coli* were lysed by French press and soluble, refolded H-2K<sup>b</sup> monomer containing a C-terminal LPETG tag followed by a 15 amino acid BirA AP tag was added.

These lysate mixtures were incubated in the presence of probe **1**, with or without sortase A for 1 hour at 37° C. SDS-PAGE resolved reactions were probed by streptavidin-HRP immunoblot and also stained with coomassie to indicate the complexity of polypeptides in the lysates. Longer exposure of streptavidin-HRP immunoblots for both (a) and (b) are shown.

**Supplementary Figure 2.5 Longer exposure of streptavidin-HRP blot from Figure 5a reveals no endogenous sortagged polypeptides.**



Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

A longer exposure of the cell surface labeling experiment with biotinylated probe **1** from **Figure 2.5a** is shown.

**Supplementary Figure 2.6 Sortagging LPETG-bearing influenza A/WSN/33 neuraminidase on the surface of live cells.**

LPETG-fused neuraminidase as well as LPETG-fused CD154 were transfected into HEK 293T cells. Cells were incubated in serum-containing media and 500  $\mu$ M probe **1**, either with or without 200  $\mu$ M sortase A for 20 minutes and collected for streptavidin-HRP immunoblot.

**Supplementary Figure 2.7 Synthesis and characterization of biotinylated probes 1-2.**

(a) The core structures of **1** and **2** were assembled on MBHA resin with an acid labile HMPB linker using standard solid phase peptide synthesis.

(b) The identity and purity of **1** and **2** were confirmed by electrospray LC/MS.

**Supplementary Figure 2.8 Synthesis and characterization of aryl azide probe 3.**

(a) The core structure of **3** was assembled on Rink Amide resin using standard solid phase peptide synthesis.

(b) The identity and purity of **3** were confirmed by electrospray LC/MS.

**Supplementary Figure 2.9 Synthesis and characterization of fluorescent probes 4-5.**

(a) The core structure of **4** and **5** was assembled on Rink Amide resin using standard solid phase peptide synthesis.

(b) The identity and purity of **4** and **5** were confirmed by electrospray LC/MS.

**Supplementary Table 2.1 Table of surface-exposed loop sequences in murine H-2K<sup>b</sup> and the sequences of LPETG substituted loops shown in Supplementary Fig. 2a.**

*Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells*

**Supplementary Figure 2.1**

**a**

**Sortase A**



Protein:

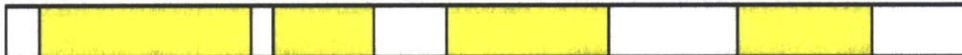
MQAKPQIPKD KSKVAGYIEI PDADIKEPVY PGPATPEQLN RGVSPAEENE  
 SLDDQNISIA GHTFIDRPNY QFTNLKAAKK GSMVYFKVGN ETRKYQMTSI  
 RDVKPTDVGV LDEQKGKDKQ LTLITCDDYN EKTGVWEKRR IPVATEVK

Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
VAGYIEIPDADIK	1403.74	8.40	14 - 26	8.78
VAGYIEIPDADIKEPVYGPATPEQLNR	3052.56	18.26	14 - 41	18.92
EPVYGPATPEQLNR	1667.84	9.98	27 - 41	10.14
KGSMVYFK	959.50	5.74	80 - 87	5.41
MTSIRDVKEPTDVGVLDEQK	2131.11	12.75	97 - 115	12.84
DVKPTDVGVLDEQK	1542.80	9.23	102 - 115	9.46
DVKPTDVGVLDEQK	1727.92	10.34	102 - 117	10.81
DKQLTITCDDYNEK	1798.85	10.76	118 - 132	10.14
QLTITCDDYNEK	1555.73	9.31	120 - 132	8.78
Totals:	8070.09	48.27	72	48.65

**b**

**H-2K<sup>b</sup>**



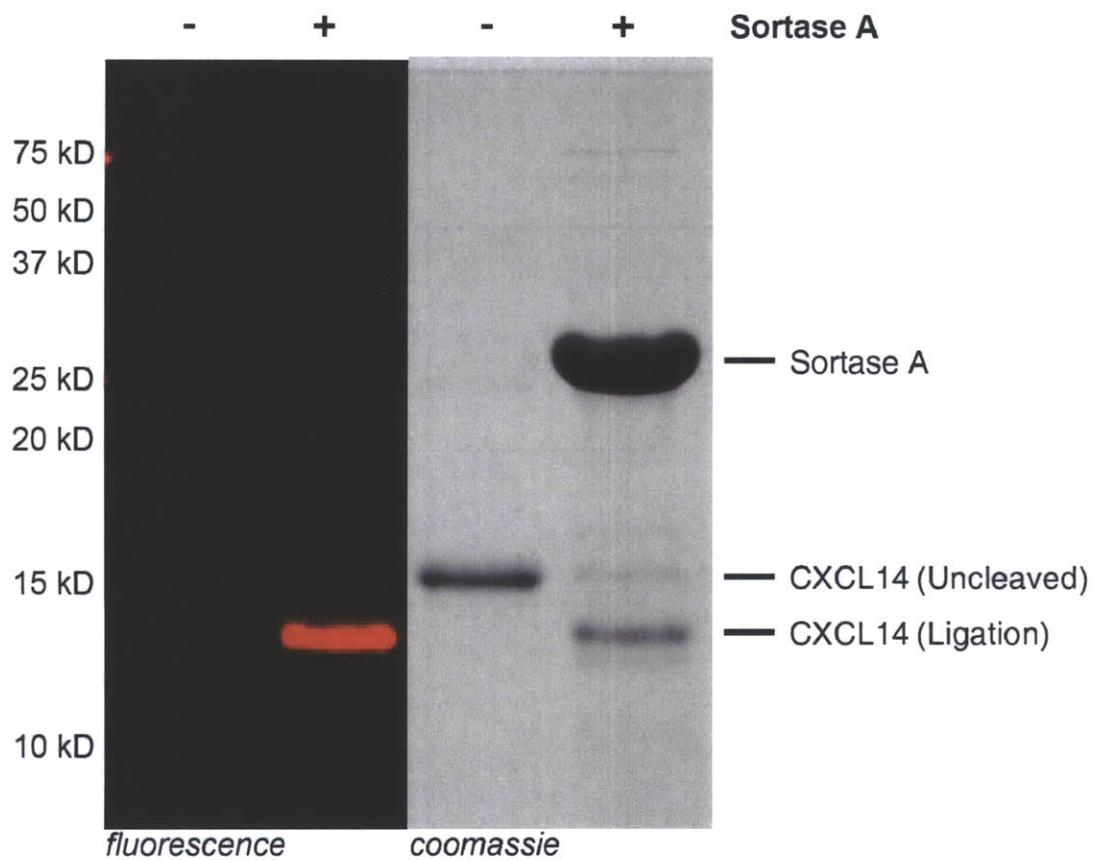
Protein:

GPHSRLRYFVT AVSRPGLGEP RYMEVGYVDD TEFVRFSDA ENPRYEPRAR  
 WMEQEGPEYW ERETQKAKGN EQSFRVLDLRT LLGYYNQSKG GSHTIQVISG  
 CEVGS DGRLL RGYQQAYDG CDYIALNEDL KTWTAADMAA LITKHEWEQA  
 GEARLRAYL EGTCEVWLR R YLK

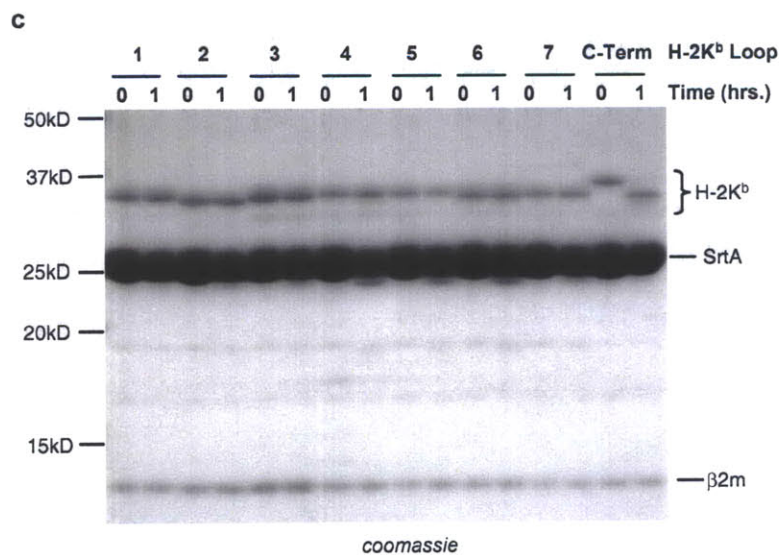
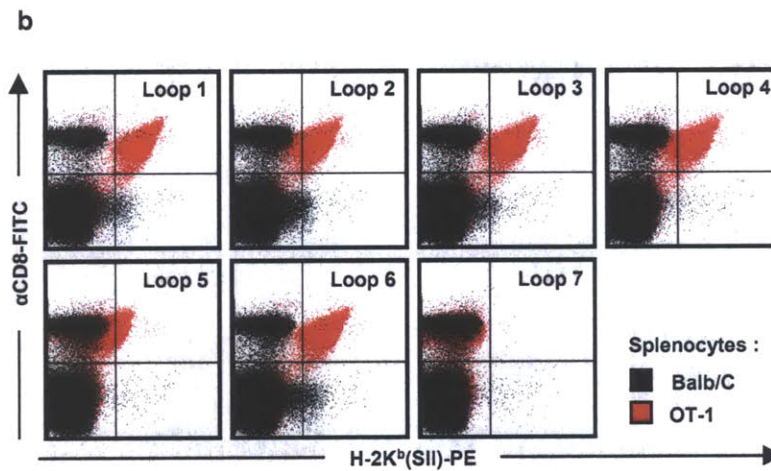
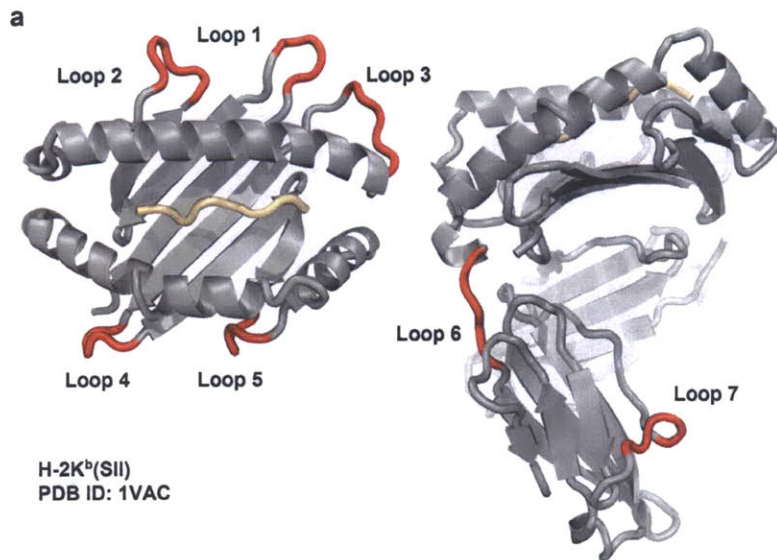
Protein Coverage:

Sequence	MH+	% Mass	AA	% AA
YFVTAISR	942.50	4.70	7 - 14	4.62
YFVTAISR PGLGEP	1648.88	8.22	7 - 21	8.67
YMEVGYVDDTEFVR	1722.77	8.58	22 - 35	8.09
FSDAENPR	1050.45	5.23	36 - 44	5.20
FSDAENPR YEPR	1595.71	7.95	36 - 48	7.51
ARWMEQEGPEYWER	1866.82	9.30	49 - 62	8.09
WMEQEGPEYWER	1639.68	8.17	51 - 62	6.94
WMEQEGPEYWERETQK	2125.93	10.59	51 - 66	9.25
ETQKAKGNEQSFR	1522.76	7.59	63 - 75	7.51
GNEQSFR	837.38	4.17	69 - 75	4.05
TLGYYNQSK	1186.61	5.91	80 - 89	5.78
GGSHTIQVISGCEVGS DGR	1858.87	9.26	90 - 108	10.98
TWTAADMAALITK	1392.72	6.94	132 - 144	7.51
HKWEQAGEAR	1340.63	6.68	145 - 155	6.36
WEQAGEAR	1075.48	5.36	147 - 155	5.20
Totals:	12420.87	61.89	109	63.01

**Supplementary Figure 2.2**

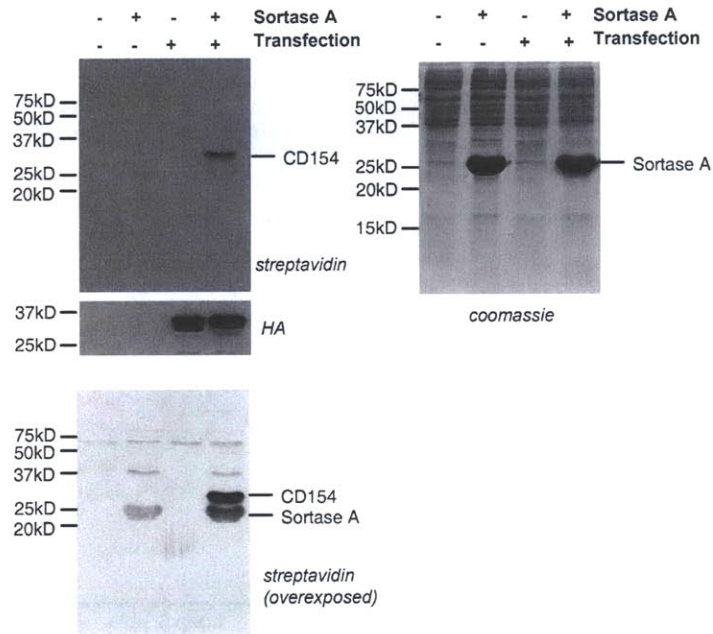


**Supplementary Figure 2.3**

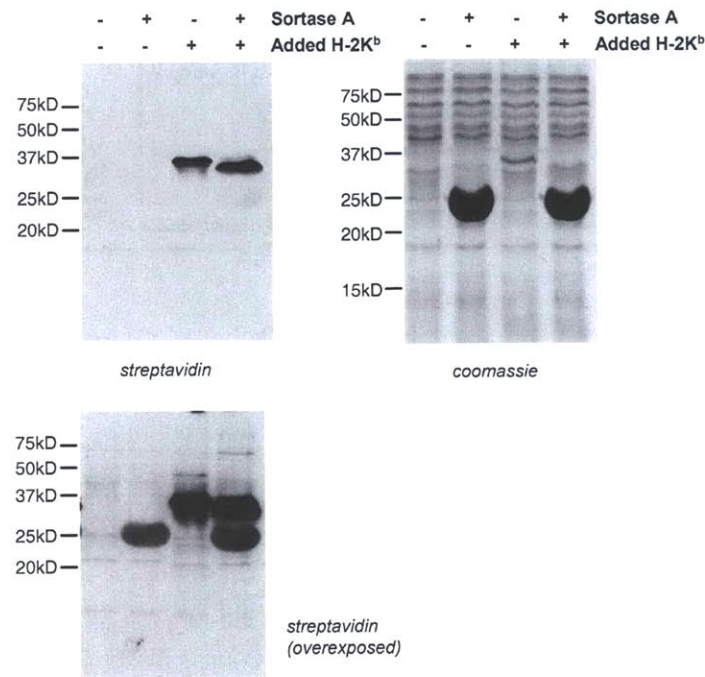


**Supplementary Figure 2.4**

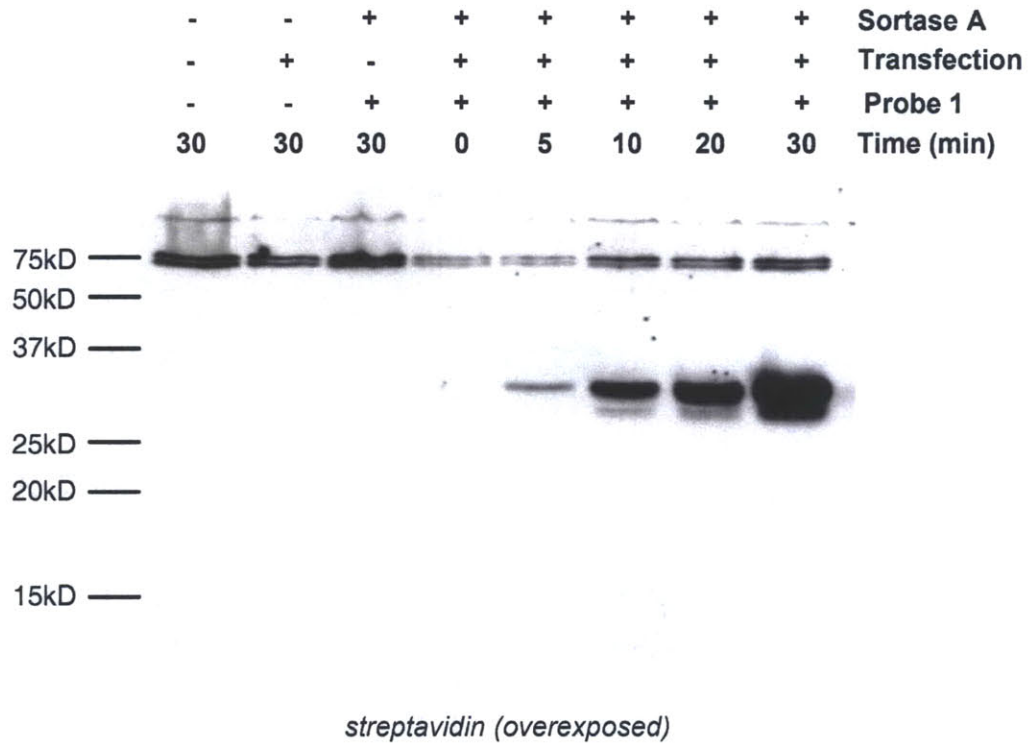
**a**



**b**

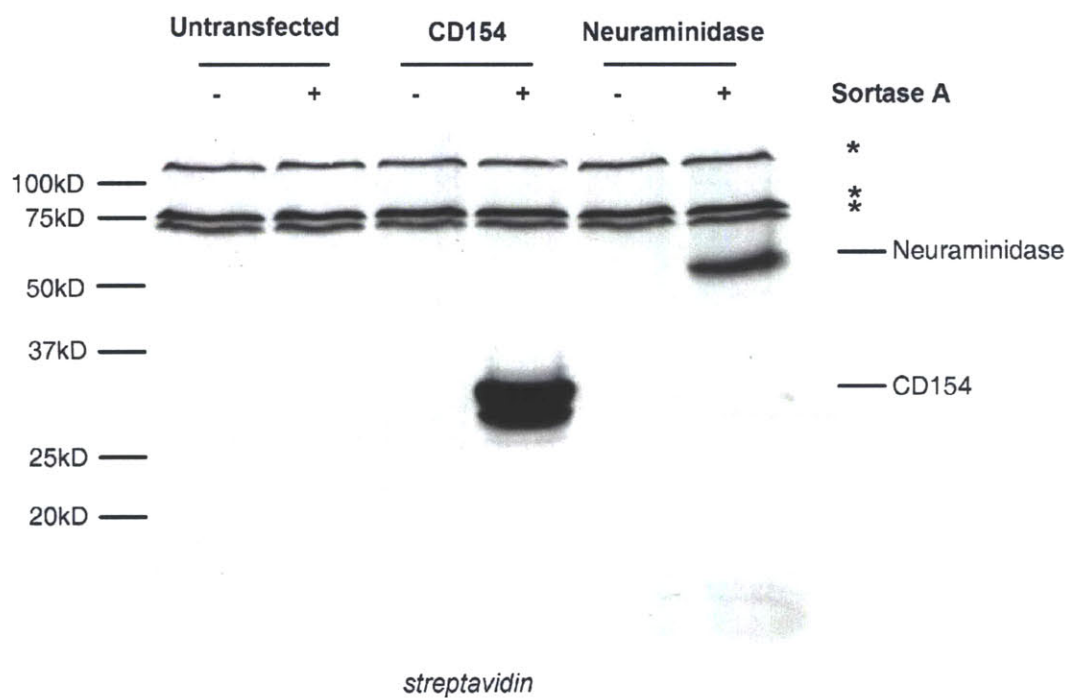


**Supplementary Figure 2.5**





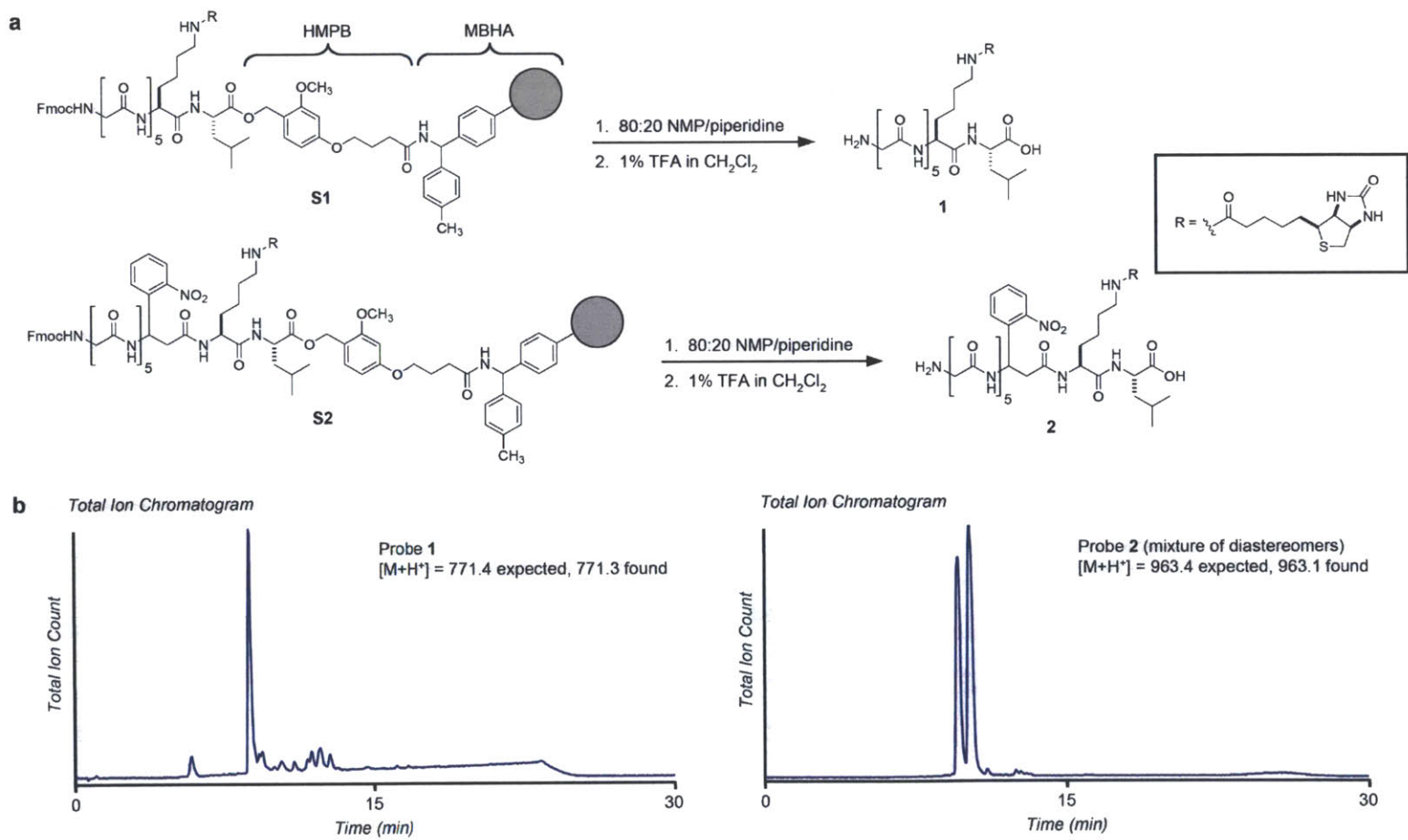
**Supplementary Figure 2.6**



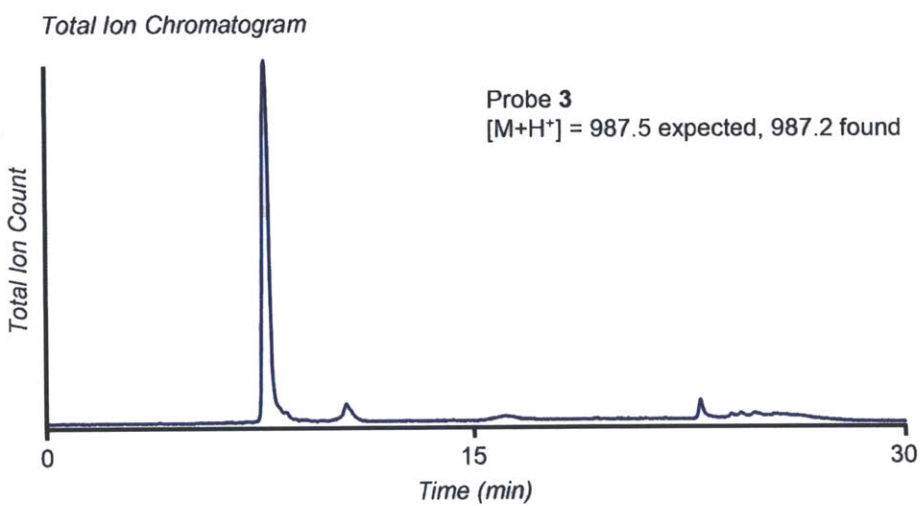
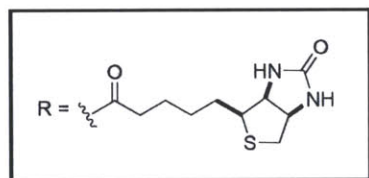
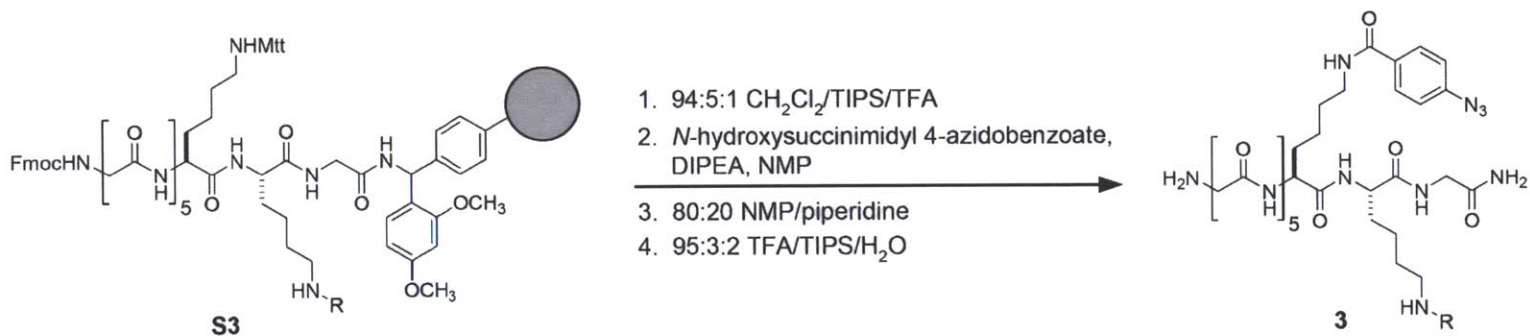
\* = Non-specific Streptavidin-HRP reactive bands



Supplementary Figure 2.7

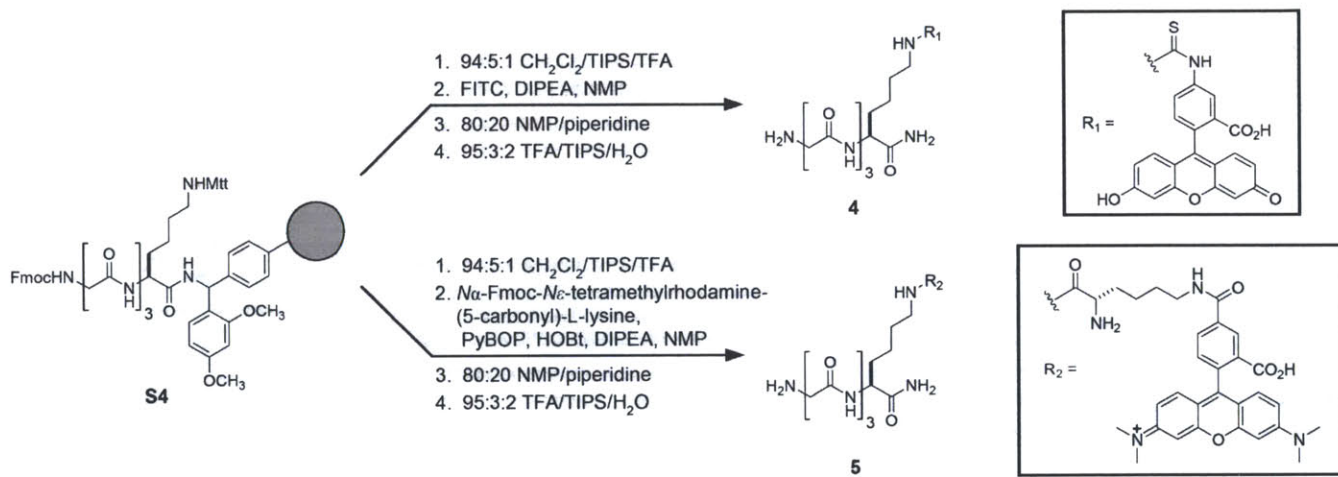


Supplementary Figure 2.8

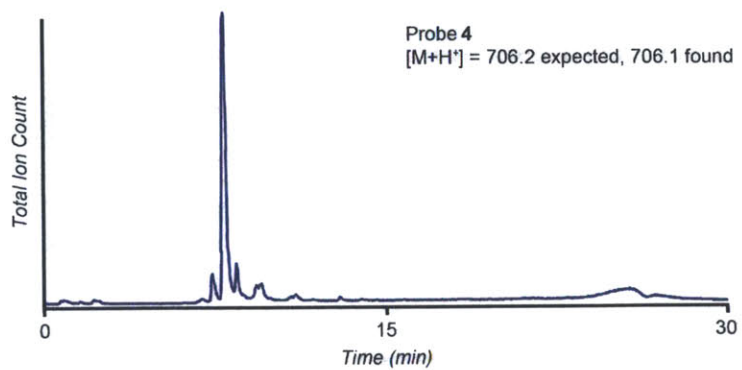


Supplementary Figure 2.9

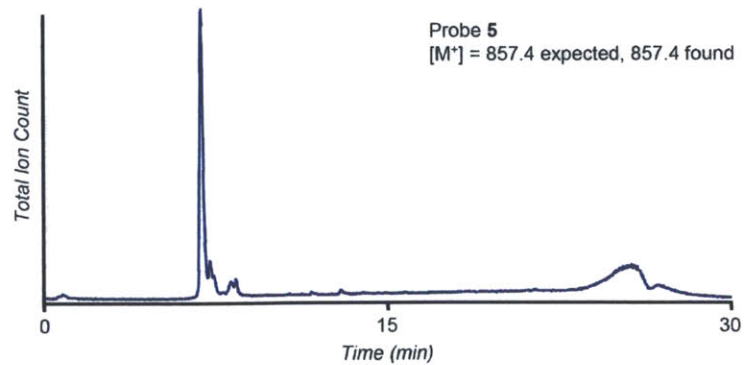
a



b Total Ion Chromatogram



Total Ion Chromatogram



**Supplementary Table 2.1**

<b>Loop</b>	<b>Residue Numbers</b>	<b>H-2K<sup>b</sup> Sequence</b>	<b>Replaced Sequence</b>
1	14-21	SRPGLGEP	SLPETGEP
2	39-50	SDAENPRYEPRA	SLPETGRYEPRA
3	87-94	NQSKGGSH	LPETGGSH
4	104-110	VGSDGRL	VLPETGL
5	127-134	LNEDLKTW	LPETGKTW
6	181-186	LRTDSP	LPETGP
7	221-235	NGEELIQDMELVETR	NGELPETGMELVETR

## Supplementary Methods

**Sortase A Expression.** Soluble *S. aureus* sortase A was expressed and purified as described<sup>14</sup>.

**Substrate Cloning and Expression.** Murine H-2K<sup>b</sup> lacking the transmembrane domain, cytosolic domain, and signal peptide, was fused to the LPETG tag followed by the BirA acceptor peptide tag (GLNDIFEAQKIEWHE) at the C-terminus and cloned into the pET28a(+) vector (Novagen). H-2K<sup>b</sup> constructs with the LPETG tag substituted into the indicated loop regions (**Supplementary Fig. 2a**) were also fused at the C-terminus to the BirA acceptor peptide and cloned into pET28a(+). The positions of the loop regions are as follows, starting from an N-terminal methionine that has been fused to the H-2K<sup>b</sup> after truncation of the signal peptide: Loop 1 = amino acids 15-19, Loop 2 = 40-44, Loop 3 = 87-91, Loop 4 = 105-109, Loop 5 = 127-131, Loop 6 = 181-185, Loop7 = 224-228 (see **Supplementary Table 1**). H-2K<sup>b</sup> constructs were expressed in BL-21 *E. coli* and soluble monomers were reconstituted in the presence of SIINFEKL peptide and  $\beta_2$ -microglobulin, as described previously<sup>29</sup>. After reconstitution, soluble H-2K<sup>b</sup> monomers were purified by size exclusion chromatography on S75 Sephadex resin using 20mM Tris, 50mM NaCl, pH 8.0 buffer as eluent.

Human CD154 (CD40L) cDNA was obtained from Open Biosystems, fused to a flexible linker of the sequence (Gly<sub>4</sub>Ser)<sub>2</sub> followed by the LPETG tag and HA epitope (YPYDVPDYA) at the extreme C-terminus and cloned into pcDNA3.1(+) (Invitrogen).

LPETG-tagged influenza A/WSN/33 was amplified and cloned similarly into pcDNA3.1(+).

**CXCL14 Substrate Cloning, Expression, and Purification.** The cDNA for human CXCL14 was obtained from the laboratory of Craig Gerard (Harvard Medical School, Boston, MA). Residues 23-99 of human CXCL14 containing an additional N-terminal glycine residue and fused to a C-terminal LPETG tag were cloned into the pET28a(+) vector (Novagen) using NcoI and XhoI restriction sites to generate a C-terminal His<sub>6</sub> affinity handle. The CXCL14 construct was expressed in BL-21 *E. coli* and isolated by Ni-NTA chromatography (Qiagen) following the general procedures described for murine CXCL14<sup>30</sup>. Refolding of the purified CXCL14 substrate was achieved as follows: A 1 mL aliquot of the eluate from Ni-affinity purification (5.5 mg/mL of CXCL14 substrate in 20 mM Tris pH 8.0 containing 6 M urea, 0.5 M NaCl, and 300 mM imidazole) was treated with 1 uL of 1 M TCEP and incubated for 1 h at RT. This material was then diluted into 100 mL of 20 mM Tris pH 8.0 and stirred rapidly at RT for 70 h. The resulting material was concentrated to ~ 1 mL by centrifugal ultrafiltration. The concentrated material was further centrifuged to pellet any precipitated protein, and the supernatant was collected. The refolded protein was analyzed by reversed-phase HPLC and eluted as a single peak with a retention time distinct from CXCL14 pretreated with excess TCEP (data not shown). No additional characterization was performed.

**CXCL14 Labeling.** Purified, refolded CXCL14 (100 μM) was labeled by incubation with 150 μM SrtA and 5 mM **5** in sortase buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

mM CaCl<sub>2</sub>) for 22 h at RT. The reaction was quenched by addition of reducing, denaturing SDS-PAGE sample buffer. Proteins were resolved on a 15% gel and visualization of fluorescent labeling was achieved with a Typhoon Imager (GE Healthcare) followed by coomassie staining of the gel (**Supplementary Figure 2**).

**Solid Phase Synthesis of Probes 1-5**

**General Procedures and Materials.** Rink Amide resin (100-200 mesh, 0.57 mmol/g), 4-methylbenzylhydramine (MBHA) resin HL hydrochloride salt (100-200 mesh, 1.1 mmol/g), 4-(4-hydroxymethyl-3-methoxyphenoxy)-butyric acid (HMPB), Fmoc-Gly-OH, Fmoc-Lys(biotin)-OH, Fmoc-Lys(Mtt)-OH, and Fmoc-Leu-OH were obtained from EMD Biosciences/Novabiochem (San Diego, CA). Racemic 3-*N*α-Fmoc-amino-3-(*o*-nitrophenyl) propionic acid (ANP linker) was purchased from Advanced ChemTech (Louisville, KY). Fluorescein isothiocyanate isomer I (FITC) and *N*-hydroxysuccinimidyl 4-azidobenzoate were obtained from Sigma. *N*α-Fmoc-*N*ε-tetramethylrhodamine-(5-carbonyl)-L-lysine was purchased from Invitrogen/Molecular Probes (Eugene, OR). All other reagents and solvents were obtained from commercial sources and used without further purification.

The identity and purity of **1-5** was confirmed by electrospray LC/MS analysis using a Micromass LCT mass spectrometer and an Agilent 1100 Series HPLC system equipped with a Waters Symmetry 3.5 μM C18 column (2.1 x 50 mm, MeCN:ddH<sub>2</sub>O gradient mobile phase containing 0.1% formic acid, 300 μL/min). HPLC purification of **2** and **3** was achieved using an Agilent 1100 Series HPLC system equipped with a Waters Delta

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

Pak 15  $\mu$ M C18 column (7.8 x 300 mm, MeCN:ddH<sub>2</sub>O gradient mobile phase containing 0.1% trifluoroacetic acid, 4 mL/min)

The peptide scaffolds (**S1-S4**, **Supplementary Figs. 4-6**) for probes **1-5** were synthesized manually using standard Fmoc-based chemistry. Unless otherwise indicated, all manipulations were conducted in glass solid-phase reaction vessels containing a fritted glass filter and a Teflon stopcock with agitation on a wrist action shaker. The following general procedure for amino acid coupling and Fmoc deprotection was employed: 2-5 equivalents (relative to initial resin loading) of Fmoc-protected amino acid was dissolved in *N*-methyl-2-pyrrolidone (NMP) to a final concentration of ~250 mM. Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) (1 equivalent relative to amino acid building block), *N*-hydroxybenzotriazole (HOBt) (1 equivalent relative to amino acid building block), and *N,N'*-diisopropylethylamine (DIPEA) (3 equivalents relative to amino acid building block) were then added and the solution was incubated for 2 min at RT. This pre-activated amino acid solution was then added to a reaction vessel containing the appropriate resin. The reaction mixture was then incubated for 3-15 h at RT. The resin was then washed with three portions of NMP (3-5 min per wash). The extent of coupling was assessed by Kaiser test. In the event that coupling was incomplete, the above procedure was repeated. Fmoc removal was then achieved by exposing the resin to 80:20 NMP/piperidine for 15 min at RT. The resin was then washed with three portions of NMP (3-5 min per wash). Additional manipulations necessary for the synthesis of **1-5**, including resin loading, fluorophore attachment, aryl azide attachment, resin cleavage, and purification are described below.



Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

**Synthesis of Biotinylated Probes 1-2.** Probes **1** and **2** were synthesized on MBHA resin employing an acid labile HMPB linker (**Supplementary Fig. 4a**). MBHA resin (1.00 g, 1.1 mmol) was first washed/swollen with 20 mL of NMP (3x, 3 min per wash). The resin was then treated with a solution of HMPB (793 mg, 3.30 mmol), PyBOP (1.72 g, 3.31 mmol), HOBt (446 mg, 3.30 mmol) and DIPEA (1.14 mL, 6.62 mmol) in NMP (15 mL) and incubated overnight at RT. The resin was then washed with 20 mL of NMP (3x, 3 min per wash) followed by 20 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 3 min per wash). Next, the resin was transferred to a round bottom flask and co-evaporated three times with 20 mL of dichloroethane (DCE). Attachment of the C-terminal Leu residue was then achieved by treating the resin with a solution of Fmoc-Leu-OH (1.17 g, 3.31 mmol), *N,N'*-diisopropylcarbodiimide (DIC) (564  $\mu$ L, 3.64 mmol) and 4-dimethylaminopyridine (DMAP) (22 mg, 0.18 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (15 mL) for two hours at RT. The resin was then filtered and washed with 20 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 3 min per wash). The coupling of Fmoc-Leu-OH was then repeated to achieve maximum resin loading. The resin was then thoroughly dried and loading was estimated to be 0.5 mmol/g by standard spectrophotometric Fmoc quantification. The dried resin was then transferred to solid-phase reaction vessels and standard cycles of Fmoc removal and amino acid coupling were used to assemble resin bound intermediates **S1** and **S2**. Removal of the N-terminal Fmoc group was then achieved by exposing the resin to 80:20 NMP/piperidine for 15 min at RT. The resin was then washed with NMP (3x, 3-5 min per wash) and CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash). Probes **1** and **2** were cleaved from the solid support by treatment with 1% TFA in CH<sub>2</sub>Cl<sub>2</sub> at RT (3x, 1<sup>st</sup> cleavage 15 min, 2<sup>nd</sup> cleavage 1.5 h, 3<sup>rd</sup> cleavage 15 min).

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

The combined cleavage solutions were then mixed with an equal volume of toluene and concentrated. The crude residues were then redissolved in 1:1 CH<sub>2</sub>Cl<sub>2</sub>/toluene and evaporated (3x) to remove excess TFA. Crude **1** and **2** were then precipitated from cold diethyl ether. Probe **2** was further purified by reversed-phase HPLC while probe **1** was used without additional purification. The identity and purity of **1** and **2** were confirmed by electrospray LC/MS (**Supplementary Fig. 4b**). Yield not determined. [Note: **2** was obtained as a mixture of diastereomers resulting from the use of racemic ANP-linker. This mixture was used for all subsequent experiments.]

**Synthesis of Aryl Azide Probe 3.** Probe **3** was synthesized on Rink Amide resin (**Supplementary Fig. 5a**). Rink Amide resin (1.00 g, 0.57 mmol) was first washed/swollen extensively with NMP. The Fmoc protecting group was then removed by treatment with 80:20 NMP/piperidine (30 mL) for 15 min at RT. The resin was then washed with 20 mL of NMP (3x, 3-5 min per wash). A solution of Fmoc-Gly-OH (847 mg, 2.85 mmol), PyBOP (1.48 g, 2.84 mmol), HOBt (385 mg, 2.85 mmol) and DIPEA (1.47 mL, 8.53 mmol) in NMP (11 mL) was then prepared and incubated for 2 min at RT. This pre-activated amino acid solution was then added to the deprotected Rink Amide resin and the mixture was incubated for 16 h at RT. The resin was then washed with 30 mL of NMP (3x, 3-5 min per wash). Kaiser test revealed the presence of free amino groups, so the coupling of Fmoc-Gly-OH was repeated. The resin was then washed with 30 mL of NMP (3x, 3-5 min per wash) and 30 mL of CH<sub>2</sub>Cl<sub>2</sub> (5x, 5 min each for first four washes, 12 h for last wash). The resin was then thoroughly dried and loading was estimated to be 0.4 mmol/g by standard spectrophotometric Fmoc quantification. The

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

dried resin was then transferred to a solid-phase reaction vessel and standard cycles of Fmoc removal and amino acid coupling were used to assemble resin bound intermediate **S3**. Resin **S3** was then washed with CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash) and dried. Dry **S3** resin (50 mg) was then transferred to a 3.0 mL fritted polypropylene syringes equipped with a capped hypodermic needle. Next, the 4-methyltrityl (Mtt) protecting group was removed by treatment with 2 mL of 94:5:1 CH<sub>2</sub>Cl<sub>2</sub>/TIPS/TFA at RT (5x, 3-5 min each). The resin was then washed with 2 mL of NMP (3x, 3-5 min per wash). Attachment of the aryl azide moiety was achieved by incubation with a solution of *N*-hydroxysuccinimidyl 4-azidobenzoate (17 mg, 65 μmol) and DIPEA (33.3 μL, 193 μmol) in NMP (0.5 mL) for 60 h at RT. The resin was then washed with 2 mL of NMP (3x, 3-5 min per wash). The N-terminal Fmoc group was then removed with 2 mL of 80:20 NMP/piperidine for 15 min at RT followed by washing with 2 mL of NMP (3x, 3-5 min per wash) and 2 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash). Probe **3** was then cleaved from the solid support with 2 mL of 95:3:2 TFA/TIPS/H<sub>2</sub>O at RT (3x, 1<sup>st</sup> cleavage 1 h, 2<sup>nd</sup> and 3<sup>rd</sup> cleavage 10 min). The combined cleavage solutions were concentrated and then co-evaporated two additional times with toluene to remove excess TFA. Crude **3** was then dried under vacuum, precipitated from cold diethyl ether, and purified by reversed-phase HPLC. The identity and purity of **3** were confirmed by electrospray LC/MS (**Supplementary Fig. 5b**). Yield not determined.

**Synthesis of Fluorescent Probes 4-5.** Probes **4** and **5** were synthesized on Rink Amide resin (**Supplementary Fig. 6a**). Rink Amide resin (1.00 g, 0.57 mmol) was first washed/swollen extensively with NMP. Standard cycles of Fmoc removal and amino

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

acid coupling were then used to assemble resin bound intermediate **S4** (resin loading was assumed to be the value given by the manufacturer, 0.57 mmol/g). Resin **S4** was then washed with CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash) and dried. Dry **S4** resin (50 mg per reaction) was then transferred to two separate 3.0 mL fritted polypropylene syringes equipped with capped hypodermic needles for fluorophore attachment. Next, the 4-methyltrityl (Mtt) protecting group was removed by treatment with 2 mL of 94:5:1 CH<sub>2</sub>Cl<sub>2</sub>/TIPS/TFA at RT (4x, 3-5 min each). The resin was then washed with 2 mL of NMP (3x, 3-5 min per wash). For FITC attachment (probe **4**), the resin was incubated with a solution of fluorescein isothiocyanate isomer I (23 mg, 59 μmol) and DIPEA (31 μL, 180 μmol) in NMP (0.4 mL) for 21 h at RT. For tetramethylrhodamine attachment (probe **5**), the resin was incubated with a solution of *N*α-Fmoc-*N*ε-tetramethylrhodamine-(5-carbonyl)-L-lysine (25 mg, 32 μmol), PyBOP (17 mg, 33 μmol), HOBt (4.3 mg, 32 μmol), and DIPEA (16 μL, 93 μmol) in NMP (0.2 mL) for 21 h at RT. Following fluorophore coupling, resins were washed with 2 mL of NMP (4x, 3-5 min per wash). The N-terminal Fmoc group was then removed with 2 mL of 80:20 NMP/piperidine for 15 min at RT followed by washing with 2 mL of NMP (4x, 3-5 min per wash) and 2 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash). Probes **4** and **5** were cleaved from the solid support with 2 mL of 95:3:2 TFA/TIPS/H<sub>2</sub>O at RT (3x, 1<sup>st</sup> cleavage 1 h, 2<sup>nd</sup> and 3<sup>rd</sup> cleavage 10 min each). The combined cleavage solutions were then concentrated and crude **4** and **5** were precipitated from cold diethyl ether. Precipitation provided **4** and **5** with sufficient purity for transpeptidation as indicated by electrospray LC/MS (**Supplementary Fig. 6b**). Yield not determined.

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

**Immunoblotting.** Proteins were separated by SDS-PAGE (12.5% gel), and transferred to a nitrocellulose membrane. Membranes were blocked with 5% nonfat dried milk in PBST (PBS, 0.1% Tween 20, pH 7.4) overnight at 4° C or for 1 hour at room temperature. Membranes were washed with PBST and incubated with either streptavidin-HRP (1:3000 in PBST, Amersham Biosciences) for 1 hour to detect biotinylated proteins, or murine anti-HA antibody (12CA5, 1:4000) followed by HRP-conjugated anti-mouse antibody (1:5000 in 5% milk, GE Healthcare). Blots were developed with Western Lighting Chemiluminescence Reagent Plus (Perkin Elmer).

**Cell Culture and Transfection.** HEK 293T cells were cultured in DME media supplemented with 10% inactivated fetal serum (IFS), 50 units/mL penicillin, 50 µg/mL streptomycin sulfate, and 0.125 µg/mL amphotericin B (Fungizone). Cells were incubated in a 5% CO<sub>2</sub> humidified incubator at 37° C. For cell surface labeling and mammalian lysate labeling with probe 1, HEK 293T cells were transfected with CD154-(Gly<sub>4</sub>Ser)<sub>2</sub>-LPETG-HA plasmid using Lipofectamine 2000 (Invitrogen) following the manufacturer's directions. For cell surface labeling with fluorescent probe 5 and subsequent microscopy, cells were transfected using Lipofectamine 2000 with a mixture of 9/10 CD154-(Gly<sub>4</sub>Ser)<sub>2</sub>-LPETG-HA plasmid and 1/10 plasmid containing soluble enhanced GFP cloned into pcDNA 3.1(+).

**H-2K<sup>b</sup> Acyl-enzyme Identification.** Labeling reactions in the absence of probe were resolved by SDS-PAGE and silver-stained. The high molecular weight band indicated (**Fig. 2A**) was excised, destained, and subjected to trypsinolysis. Recovered peptides

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

were analyzed by reversed-phase liquid chromatography electrospray ionization mass spectrometry using a Waters nanoACQUITY-UPLC coupled to a Thermo LTQ linear ion-trap mass spectrometer. MS/MS spectra were searched against NCBI database (nr.fasta.hr 6/27/2006) using SEQUEST. SEQUEST results were analyzed with Bioworks Browser 3.2 and filtered with the following criteria: different peptides, minimum cross correlation coefficients (1, 2, 3 charge states) of 1.50, 2.00, 2.50, number different peptides of 2 per protein and Sp – preliminary score of 300.

**HEK 293T Lysate Labeling.** Approximately 12-24 hours after transfection, cells were harvested into ice-cold PBS by scraping and pelleted. Cell pellets were lysed in ice-cold 0.5 % NP-40 in sortase buffer and protein concentration was determined by BCA assay (Pierce). SrtA (5  $\mu$ g) was mixed with lysate (10  $\mu$ g) and biotinylated probe 1 (100  $\mu$ M), supplemented with sortase buffer to give a final concentration of 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub> and incubated for 1 hour at 37° C. Reactions were halted by addition of reducing, denaturing SDS-PAGE sample buffer.

**Bacterial Lysate Labeling.** BL-21 *E. coli* were resuspended in sortase buffer and lysed by French Press. Protein concentration was determined by BCA assay. Lysate (10  $\mu$ g) was spiked with soluble, refolded sortagged H-2K<sup>b</sup> monomer (2.24  $\mu$ g) bearing BirA AP, incubated with SrtA (10  $\mu$ g) and biotinylated probe 1 (500  $\mu$ M), and supplemented with sortase buffer to give a final concentration of 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub> for 1 hour at 37° C. Reactions were halted by addition of reducing, denaturing

Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

SDS-PAGE sample buffer. For these reactions the BirA AP region of H-2K<sup>b</sup> was not additionally biotinylated with BirA prior to sortagging.

**Supplementary Methods References**

1. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci U S A* **96**, 12424-9 (1999).
2. Garboczi, D. N., Hung, D. T. & Wiley, D. C. HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in *Escherichia coli* and complexed with single antigenic peptides. *Proc Natl Acad Sci U S A* **89**, 3429-33 (1992).
3. Sleeman, M. A. et al. B cell- and monocyte-activating chemokine (BMAC), a novel non-ELR {alpha}-chemokine 10.1093/intimm/12.5.677. *Int. Immunol.* **12**, 677-689 (2000).

Print Article (reprinted with permission from Nature Chemical Biology):

BRIEF COMMUNICATIONS

nature  
chemical biology

## Sortagging: a versatile method for protein labeling

Maximilian W Popp<sup>1,2</sup>, John M Antos<sup>1</sup>, Gijsbert M Grotenbreg<sup>1</sup>, Eric Spooner<sup>1</sup> & Hidde L Ploegh<sup>1,2</sup>

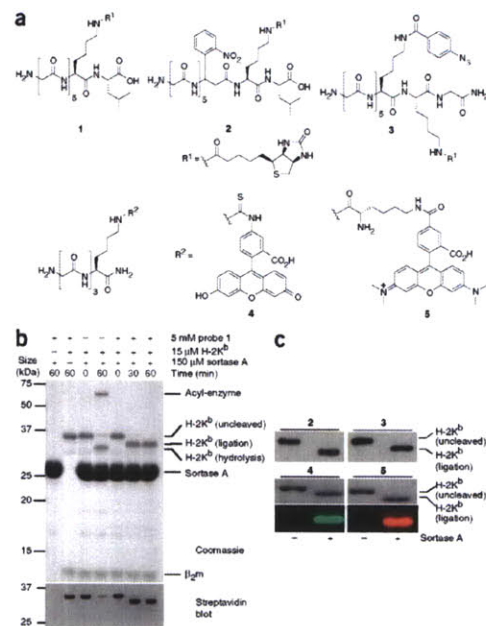
**Genetically encoded reporter constructs that yield fluorescently labeled fusion proteins are a powerful tool for observing cell biological phenomena, but they have limitations. Sortagging (sortase-mediated transpeptidation) is a versatile chemoenzymatic system for site-specific labeling of proteins with small (<2 kDa) probes. Sortagging combines the precision of a genetically encoded tag with the specificity of an enzymatic reaction and the ease and chemical versatility of peptide synthesis. Here we apply this technique to proteins *in vitro* and on the surface of living cells.**

The use of green fluorescent protein (GFP) and its derivatives has revolutionized the study of protein behavior in living cells<sup>1</sup>. However, not all proteins tolerate the installation of GFP without compromising function or intracellular distribution<sup>2</sup>. Chemical methods for the installation of fluorescent or affinity labels have the advantage of ease of use, but they lack the precision of genetically encoded tags. To overcome this challenge, chemoenzymatic methods and small-molecule-binding peptide sequences that allow site-specific incorporation of labels have been developed. However, several of these methods are limited by the large protein-sized fusions required or the demands of synthetic skills involved in producing the reporter molecules that can be installed<sup>3,4</sup> (reviewed in ref. 5).

Bacterial sortases are thiol-containing enzymes that covalently attach proteins to the bacterial cell wall<sup>6</sup>. *Staphylococcus aureus* sortase A recognizes a set of structurally and functionally diverse substrates via an LPXTG motif and cleaves the peptide bond between threonine and glycine, thereby releasing the residues C-terminal to the threonine and yielding an amide linkage with the N terminus of a pentaglycine nucleophile, which is provided *in vivo* by a cell wall precursor<sup>7</sup>. Sortase A has been used *in vitro* to affix cell-permeable peptides<sup>8</sup>, PEG and polystyrene beads<sup>9</sup> to a single model substrate, GFP; it has also been

used in the synthesis of peptide nucleic acid–peptide conjugates<sup>10</sup>. However, the range of applications was limited and did not include intact cells. Here we exploit the transpeptidase activity of sortase A for selective labeling of proteins in solution, in cell lysates of complex composition and on the surface of living cells with a diverse set of readily synthesized probes suitable for the study of protein interactions and protein trafficking (Fig. 1a).

As a substrate, we used soluble mouse class I major histocompatibility complex (MHC) H-2K<sup>b</sup> molecules equipped with an LPETG motif followed by a C-terminal 15 amino acid acceptor peptide sequence for *Escherichia coli* BirA biotin ligase and loaded with the octapeptide ligand SIINFEKL (ref. 11). We biotinylated the acceptor peptide portion and incubated the LPETG-containing H-2K<sup>b</sup> monomers (15 μM) with sortase A (150 μM) in the presence or absence of biotinylated pentaglycine probe 1 (5 mM; Fig. 1b and Supplementary Methods online). In the absence of probe 1 we observed loss of biotin, which corresponds to departure of the biotinylated acceptor peptide



**Figure 1** Probes compatible with sortagging and quantitative labeling of H-2K<sup>b</sup> complexes *in vitro*. (a) Structures of probes used. (b) H-2K<sup>b</sup> monomers with a C-terminal LPETG tag followed by a BirA acceptor peptide were incubated with sortase A in the presence or absence of probe 1. See text for details. (c) Probes 2 and 3 are quantitatively appended to H-2K<sup>b</sup> as assessed by streptavidin-HRP immunoblot, as are probes 4 and 5, as assessed by Coomassie staining (top) and fluorescence imaging (bottom). β2m, β<sub>2</sub>-microglobulin.

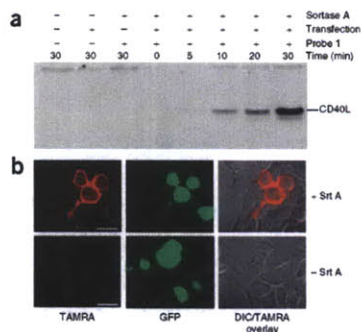
<sup>1</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02142, USA. Correspondence should be addressed to H.L.P. (ploegh@wi.mit.edu).

Received 2 April; accepted 30 July; published online 23 September 2007; doi:10.1038/nchembio.2007.31

NATURE CHEMICAL BIOLOGY | ADVANCE ONLINE PUBLICATION



BRIEF COMMUNICATIONS



**Figure 2** Labeling of proteins on cell surfaces. (a) HEK 293T cells expressing CD40L fused to an LPETG motif followed by a C-terminal HA epitope were incubated with sortase A and probe 1 and analyzed by streptavidin-HRP immunoblot. (b) Cells expressing soluble, cytoplasmic eGFP and LPETG-fused CD40L were incubated with sortase A and probe 5 for 10 min, fixed and imaged. Scale bars, 10  $\mu$ M.

tag and the concomitant formation of the acyl-enzyme intermediate<sup>12</sup>, as determined by mass spectrometry (Supplementary Fig. 1a,b online). When we added probe 1 and drove the reaction to completion by excess sortase A, we observed the expected transpeptidation product: a sortase-cleaved, biotinylated H-2K<sup>b</sup> molecule (Fig. 1b).

Neither the LPETG tag nor the transpeptidation reaction perturbs the fold of this complex, as verified by fluorescence-activated cell sorting (FACS) staining of OT-I T cells with tetramerized, sortagged H-2K<sup>b</sup> molecules (Supplementary Fig. 1c). We investigated the requirements for LPETG tag placement within the defined structure of the H-2K<sup>b</sup> protein by preparing H-2K<sup>b</sup> substrates in which the LPETG motif was moved to seven surface-exposed loop regions (Supplementary Fig. 2a and Supplementary Table 1 online). Proper folding of these substrates was largely retained, as indicated by their ability to stain OT-I T cells after tetramerization (Supplementary Fig. 2b). However, on exposure to sortase A and probe 1, no transpeptidation product was detected (Supplementary Fig. 2c). Thus, we conclude that the LPETG tag must be placed in a flexible, unstructured region close to the C terminus of the substrate.

We synthesized additional oligoglycine nucleophiles (probes 2, 3, 4 and 5) and demonstrated their ability to efficiently label the H-2K<sup>b</sup> substrate (Fig. 1c). Introduction of a 3-amino-3-(*o*-nitrophenyl)propionic acid residue (probe 2) allowed photocleavage of the tag under mild conditions, with concomitant release of the biotin label (Supplementary Fig. 3a online). Similarly, an aryl azide photocross-linker (probe 3) was sufficient to cross-link the sortagged H-2K<sup>b</sup> heavy chain to  $\beta_2$ -microglobulin (Supplementary Fig. 3b). Finally, a fluorescein (FITC) or tetramethylrhodamine (TAMRA) dye (probes 4 and 5, respectively) allowed fluorescent visualization. In addition to H-2K<sup>b</sup>, we have also successfully labeled an LPETG-tagged version of the chemokine CXCL14 in solution (Supplementary Fig. 3c).

We investigated the specificity of the transpeptidation reaction in complex mixtures. Human CD154 (CD40L), a type II membrane protein, was equipped with a C-terminal LPETG tag preceded by a short flexible linker and followed by the HA epitope tag. Confirming the specificity of the method, CD40L and H-2K<sup>b</sup> were the only species labeled in transfected human embryonic kidney (HEK) 293T cell

lysates (Supplementary Fig. 4a online) and bacterial lysates (Supplementary Fig. 4b), respectively.

Next, live HEK 293T cells transfected with the CD40L construct were incubated with serum-containing medium and sortase A together with probe 1. Streptavidin-HRP immunoblotting revealed selective labeling of CD40L with almost no labeling of endogenous polypeptides (Fig. 2a). We detected the persistence of HA-tagged CD40L, which indicates that sortase A does not attack the entire pool of CD40L substrate, possibly because not all CD40L is expressed at the cell surface, or because only a portion is accessible to sortase (Supplementary Fig. 5a online). HEK 293T cells cotransfected with the CD40L substrate and soluble eGFP were labeled with the TAMRA-containing probe 5 after just 10 min (Fig. 2b). We have also labeled surface-displayed influenza A/WSN/33 neuraminidase by sortagging (Supplementary Fig. 5b).

Recombinant His<sub>6</sub>-tagged sortase A can be produced in good yield and purity, and because of the transpeptidation reaction mechanism, may be simultaneously removed from crude reaction mixtures along with unreacted substrate by including a His<sub>6</sub> tag C-terminal to the requisite LPXTG motif. When used on intact cells, simple washing effectively removes sortase A. In addition, oligoglycine probes are accessible by standard solid-phase methods for peptide synthesis (Supplementary Fig. 6 online). Sortagging derives its utility from a genetically encoded tag of five residues<sup>13,14</sup> preceded by a short spacer and is suitable for labeling purified and surface-displayed proteins with a suitably exposed LPETG motif close to the C terminus. The multiplicity of sortases and their distinct recognition sequences<sup>15</sup> in principle immediately expands the range of labeling possibilities to the simultaneous use of several uniquely tagged proteins and various probes in a single experimental setting.

Note: Supplementary information and chemical compound information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS

We thank O. Schneewind for the sortase A-encoding plasmid, C.H. Shen and J. Chen for the neuraminidase template, and the Ploegh lab and H. Hang for helpful discussions. This work was supported by grants from the US National Institutes of Health.

AUTHOR CONTRIBUTIONS

M.W.P., J.M.A. and G.M.G. performed experiments and prepared the manuscript; E.S. performed mass spectrometry analysis; H.L.P. prepared the manuscript and conceived of the project.

Published online at <http://www.nature.com/naturechemicalbiology>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- Lippincott-Schwartz, J. & Patterson, G.H. *Science* **300**, 87–91 (2003).
- Lisenbee, C.S., Kamik, S.K. & Trelease, R.N. *Traffic* **4**, 491–501 (2003).
- Galetta, G. *et al. Science* **296**, 503–507 (2002).
- Roberti, M.J., Bartocchini, C.W., Klement, R., Jares-Erijman, E.A. & Jovin, T.M. *Nat. Methods* **4**, 345–351 (2007).
- Chen, I. & Ting, A.Y. *Curr. Opin. Biotechnol.* **16**, 35–40 (2005).
- Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. *Proc. Natl. Acad. Sci. USA* **96**, 12424–12429 (1999).
- Marraffini, L.A. & Schneewind, O. *Mol. Microbiol.* **62**, 1402–1417 (2006).
- Mao, H., Hart, S.A., Schink, A. & Pollak, B.A. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).
- Parthasarathy, R., Subramanian, S. & Boder, E.T. *Bioconjug. Chem.* **18**, 469–476 (2007).
- Pritz, S. *et al. J. Org. Chem.* **72**, 3909–3912 (2007).
- Skinner, P.J., Daniels, M.A., Schmidt, C.S., Jameson, S.C. & Haase, A.T. *J. Immunol.* **165**, 613–617 (2000).
- Aulabaugh, A. *et al. Anal. Biochem.* **360**, 14–22 (2007).
- Kruger, R.G. *et al. Biochemistry* **43**, 1541–1551 (2004).
- Zong, Y., Bice, T.W., Ton-That, H., Schneewind, O. & Narayana, S.V. *J. Biol. Chem.* **279**, 31383–31389 (2004).
- Drams, S., Trieu-Cuot, P. & Bierné, H. *Res. Microbiol.* **156**, 289–297 (2005).



Chapter 2: Sortagging (sortase-mediated transpeptidation): A versatile method for site-specific labeling of proteins in solution and on live cells

**Supplementary Online Material at:**

**[http://www.nature.com/nchembio/journal/v3/n11/supinfo/nchembio.2007.31\\_S1.html](http://www.nature.com/nchembio/journal/v3/n11/supinfo/nchembio.2007.31_S1.html)**

*Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities*

**Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities**

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

**Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities**

(from J. M. Antos, G. L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W. Popp, H. L. Ploegh, *J Am Chem Soc* **2009**, *131*, 10800)

**Abstract**

The unique reactivity of two sortase enzymes, SrtA<sub>staph</sub> from *Staphylococcus aureus* and SrtA<sub>strep</sub> from *Streptococcus pyogenes*, is exploited for site-specific labeling of a single polypeptide with different labels at its N and C termini. SrtA<sub>strep</sub> is used to label the protein's C terminus at an LPXTG site with a fluorescently labeled dialanine nucleophile. Selective N-terminal labeling of proteins containing N-terminal glycine residues is achieved using SrtA<sub>staph</sub> and LPXT derivatives. The generality of N-terminal labeling with SrtA<sub>staph</sub> is demonstrated by near-quantitative labeling of multiple protein substrates with excellent site specificity.

**Main Text**

Methods for site-specific modification of proteins remain in high demand. The transpeptidation reaction catalyzed by sortase A from *Staphylococcus aureus* (SrtA<sub>staph</sub>) allows site-specific derivatization of proteins with virtually any type of functional material.<sup>(1)</sup> Target proteins are engineered to contain the SrtA<sub>staph</sub> recognition site (LPXTG) near their C terminus, thus allowing a transacylation reaction in which the residues C-terminal to threonine are exchanged for a synthetic oligoglycine peptide (**Scheme 3.1**). While the range of applications for this technology has expanded considerably<sup>1</sup> the ligation chemistry itself has seen relatively few modifications or improvements. Since nearly all Gram-positive bacteria possess sortases, many with

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

reactivity distinct from SrtA<sub>staph</sub>, there is an exciting opportunity to develop complementary strategies for protein engineering using other members of this enzyme family<sup>2,3</sup>. Here we present a strategy for placing discrete labels at both termini in the same polypeptide through the use of multiple sortases. We first describe the ability of SrtA<sub>staph</sub> to append labels at the protein N terminus and then demonstrate how this can be used in conjunction with the activity of sortase A from *Streptococcus pyogenes* (SrtA<sub>strep</sub>) to yield dual-labeled proteins.

With regard to N-terminal labeling mediated by SrtA<sub>staph</sub>, we reasoned that labeled synthetic peptides containing the LPXTG recognition motif, or structural analogues thereof, could generate the requisite acyl–enzyme intermediate necessary for transpeptidation. In combination with protein nucleophiles containing one or more N-terminal glycines, this should result in transfer of the label to the protein N terminus (**Figure 3.1a**). We synthesized FITC (**1**) and biotin (**2**) derivatives of an LPRT peptide in which the glycine of the normal LPXTG motif was replaced by a methyl ester (**Figure 3.1b** and **Supplementary Figure 3.1** in the **Supporting Information**). The use of an ester derivative rather than the entire LPXTG motif was motivated by our concern that the glycine residue released after LPXTG cleavage might compete with the protein nucleophile, potentially complicating the desired N-terminal-labeling reaction. In contrast, transacylation with **1** and **2** would generate MeOH, a poor nucleophile for transacylation compared with glycine. It should be noted that concurrently with the work described here, it was demonstrated that labeled LPETGG peptides are viable tools for N-terminal labeling using SrtA<sub>staph</sub><sup>4</sup>.

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

With **1** and **2** in hand, we expressed a series of model proteins containing N-terminal glycine residues. We prepared variants of the cholera toxin B subunit (CtxB) with one, three, or five N-terminal glycines. In order to verify the selectivity for glycine, we prepared a control construct containing an N-terminal alanine residue. In the presence of SrtA<sub>staph</sub>, we observed robust labeling of G<sub>3</sub>-CtxB and G<sub>5</sub>-CtxB using 500 μM **1** for 2 h at 37 °C, with no apparent labeling of the alanine-containing control (Figure 1c).

Electrospray ionization mass spectrometry (ESI-MS) revealed quantitative labeling of G<sub>3</sub>-CtxB and G<sub>5</sub>-CtxB, with no modification observed for G<sub>1</sub>-CtxB and AG<sub>4</sub>-CtxB (**Supplementary Figure 3.2**). Similar experiments with biotinylated derivative **2** and G<sub>5</sub>-CtxB yielded comparable results, as verified by ESI-MS and streptavidin immunoblot (**Supplementary Figure 3.3**). In all cases, residual labeling of SrtA<sub>staph</sub> itself, attributable to the formation of a covalent acyl-enzyme intermediate, was detected. N-terminal transpeptidation was also successful for two additional protein substrates, eGFP with five N-terminal glycines and UCHL3 containing a single N-terminal glycine (**Supplementary Figure 3.4**).

With the ability of SrtA<sub>staph</sub> to append labels at either terminus, we pursued the possibility of installing two modifications within the same protein. Attempts to execute this type of transformation using SrtA<sub>staph</sub> alone were unsuccessful, as intramolecular transpeptidation between N-terminal glycines and the C-terminal LPXTG motif was unavoidable in most cases. Therefore, we considered the possibility of using a second, distinct sortase, an idea that has been suggested but never reduced to practice<sup>1,5</sup>. We initially sought to use

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

sortase B (SrtB) from either *Staph. aureus* or *Bacillus anthracis* as enzymes with recognition sequences (NPQTN and NPKTG, respectively) orthogonal to that of SrtA<sub>staph</sub><sup>6,7</sup>. Both SrtB enzymes were easily produced in *Escherichia coli* and purified to homogeneity. We reproduced the reported in vitro enzyme activity using a FRET-based assay to measure cleavage of short peptides substrates<sup>6,7</sup>. However, to date we have failed to obtain transpeptidation with either SrtB on protein substrates modified with the appropriate recognition sequences on a time scale or with yields that compare favorably with SrtA<sub>staph</sub> (data not shown).

We ultimately arrived at a successful orthogonal strategy using SrtA<sub>strep</sub>, which recognizes the same LPXTG sequence used by SrtA<sub>staph</sub> but can accept alanine-based nucleophiles<sup>8</sup>. This leads to the formation of an LPXTA sequence at the site of ligation, a motif refractory to cleavage by SrtA<sub>staph</sub><sup>9</sup>. This allows SrtA<sub>staph</sub> to act on the N terminus without affecting the C-terminal modification installed with SrtA<sub>strep</sub>.

Our final strategy for dual-terminus labeling is outlined in **Figure 3.2b**. We first synthesized a tetramethylrhodamine-labeled peptide (**3**) containing two N-terminal alanine residues to serve as the nucleophile for SrtA<sub>strep</sub>-mediated protein ligation (**Figure 3.2a and Supplementary Figure 3.5**). We prepared two model substrates (eGFP and UCHL3) containing masked N-terminal glycines that are exposed only upon thrombin cleavage. Masking was required because SrtA<sub>strep</sub> was observed to ligate both glycine and alanine nucleophiles (data not shown). Substrates also contained an LPXTG motif at the

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

C terminus to allow a first round of labeling with SrtA<sub>strep</sub>. For both eGFP and UCHL3, C-terminal labeling using **3** and SrtA<sub>strep</sub> resulted in >90% conversion to the desired adduct, as revealed by ESI-MS (**Supplementary Figure 3.6**). SrtA<sub>strep</sub> was quenched by the addition of MTSET followed by removal of His<sub>6</sub>-tagged SrtA<sub>strep</sub> using Ni-NTA. Residual **3** was then removed using a disposable desalting column. Thrombin cleavage proceeded in quantitative fashion using commercial thrombin agarose resin (**Supplementary Figure 3.6**). The exposed N-terminal glycines were then labeled by treatment with 500 μM **1** and 50 μM Δ59-SrtA<sub>staph</sub><sup>10</sup> for 1 h at 37 °C. ESI-MS of crude reaction mixtures showed the dual-labeled material as the major component, with only minor amounts of byproduct (**Supplementary Figure 3.6**). A final separation by anion-exchange chromatography yielded dual-labeled eGFP and UCHL3 with excellent purity, as determined by both SDS-PAGE and ESI-MS (**Figure 3.2c,d** and **Supplementary Figure 3.6**). In the case of UCHL3, we observed some additional low-intensity bands in the fluorescent gel scan (**Figure 3.2d**). However, quantitative densitometric analysis of coomassie-stained gels indicated purity in excess of 95% for both dual-labeled eGFP and UCHL3.

In summary, we have developed a strategy for placing different chemical labels at the two ends of the same polypeptide using two sortase enzymes with unique reactivities. We anticipate that this method will be applicable to the preparation of protein conjugates for refolding studies or for the construction of protein sensors, where measuring conformational changes by FRET is a common mode of detection. In more general terms, this work begins to explore the range of protein modifications that can be accessed using



Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

alternative sortases. The number of sortases that have been produced in recombinant form with retention of activity is continually increasing, and we are exploring the use of these unique enzymes as tools for protein engineering.

**Acknowledgment**

This work was supported by grants from the National Institutes of Health (R01-AI057182, R01-AI033456, R21-EB008875). J.M.A. and N.C.Y. acknowledge Clay Postdoctoral Fellowships. C.P.G. thanks the Fundacao para a Ciencia e Tecnologia (Portugal). G.-L.C. was supported by MIT's Undergraduate Research Opportunities Program (UROP). The authors thank Mark J. Banfield and colleagues for providing the expression plasmid for SrtA<sub>strep</sub> and Olaf Schneewind for providing the SrtB constructs.

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

**References**

1. Tsukiji, S. and Nagamune, T. *ChemBioChem* **2009**, 10, 787–798
2. Marraffini, L. A., Dedent, A. C. and Schneewind, O. *Microbiol. Mol. Biol. Rev.* **2006**, 70, 192– 221
3. Pallen, M. J., Lam, A. C., Antonio, M. and Dunbar, K. *Trends Microbiol.* **2001**, 9, 97– 102
4. Yamamoto, T. and Nagamune, T. *Chem. Commun.* **2009**, 1022– 1024
5. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. and Ploegh, H. L. *Nat. Chem. Biol.* **2007**, 3, 707– 708
6. Maresso, A. W., Chapa, T. J. and Schneewind, O. *J. Bacteriol.* **2006**, 188, 8145– 8152
7. Mazmanian, S. K., Ton-That, H., Su, K. and Schneewind, O. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 2293– 2298
8. Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., Sessions, R. B., Kehoe, M. A., McCafferty, D. G. and Banfield, M. J. *J. Biol. Chem.* **2009**, 284, 6924– 6933
9. Kruger, R. G., Otvos, B., Frankel, B. A., Bentley, M., Dostal, P. and McCafferty, D. G. *Biochemistry* **2004**, 43, 1541– 1551
10.  $\Delta 59\text{-SrtA}_{\text{staph}}$  is a truncated form of  $\text{SrtA}_{\text{staph}}$  that has identical reactivity .

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

**Figure Legends**

**Scheme 3.1. C-terminal labeling using SrtA<sub>staph</sub>.**

**Figure 3.1. N-terminal labeling using SrtA<sub>staph</sub>.**

- (a) SrtA<sub>staph</sub> catalyzes a transacylation reaction using labeled LPRT methyl esters as substrates. The labeled LPRT fragment is transferred to proteins containing N-terminal glycines in a site-specific fashion.
- (b) FITC (1) and biotin (2) LPRT methyl esters for N-terminal transacylation.
- (c) CtxB derivatives (50  $\mu$ M) were treated with 500  $\mu$ M 1 and 50  $\mu$ M SrtA<sub>staph</sub> for 2 h at 37 °C in 50 mM Tris (pH 7.5), 150 mM NaCl, and 10 mM CaCl<sub>2</sub>. Reactions were analyzed by SDS-PAGE with visualization by coomassie staining and fluorescent gel scanning.

**Figure 3.2. Site-specific N- and C-terminal labeling using multiple sortases.**

- (a) Tetramethylrhodamine-labeled dialanine nucleophile (3) for SrtA<sub>strep</sub>-mediated transpeptidation.
- (b) Strategy for the installation of discrete labels at both termini of the same protein using  $\Delta$ 59-SrtA<sub>staph</sub> and SrtA<sub>strep</sub>.
- (c, d) SDS-PAGE characterization with fluorescent gel scanning of dual-labeled (c) eGFP and (d) UCHL3.

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

Scheme 3.1

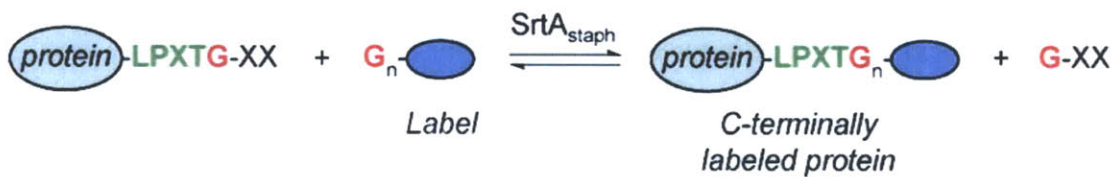
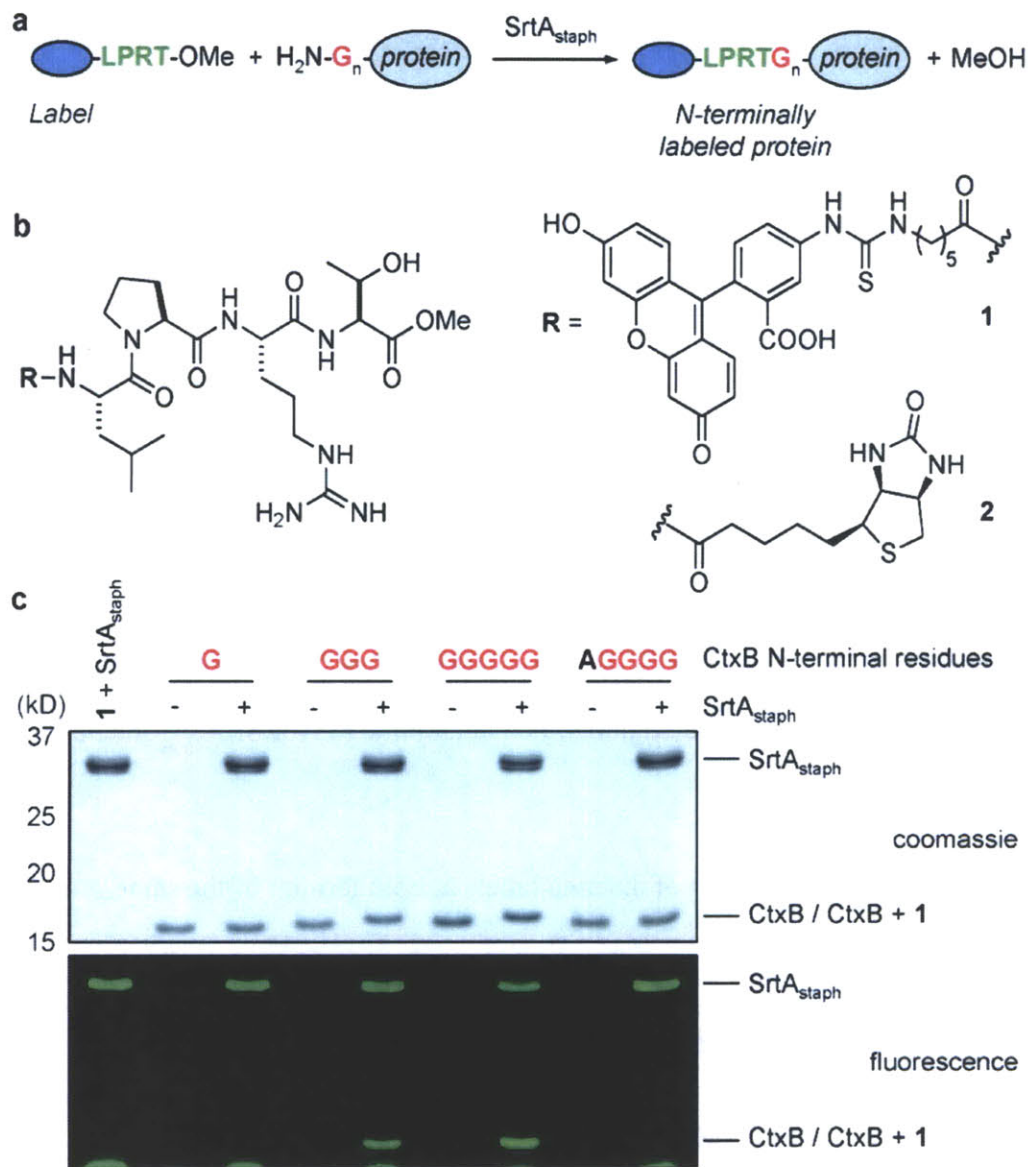
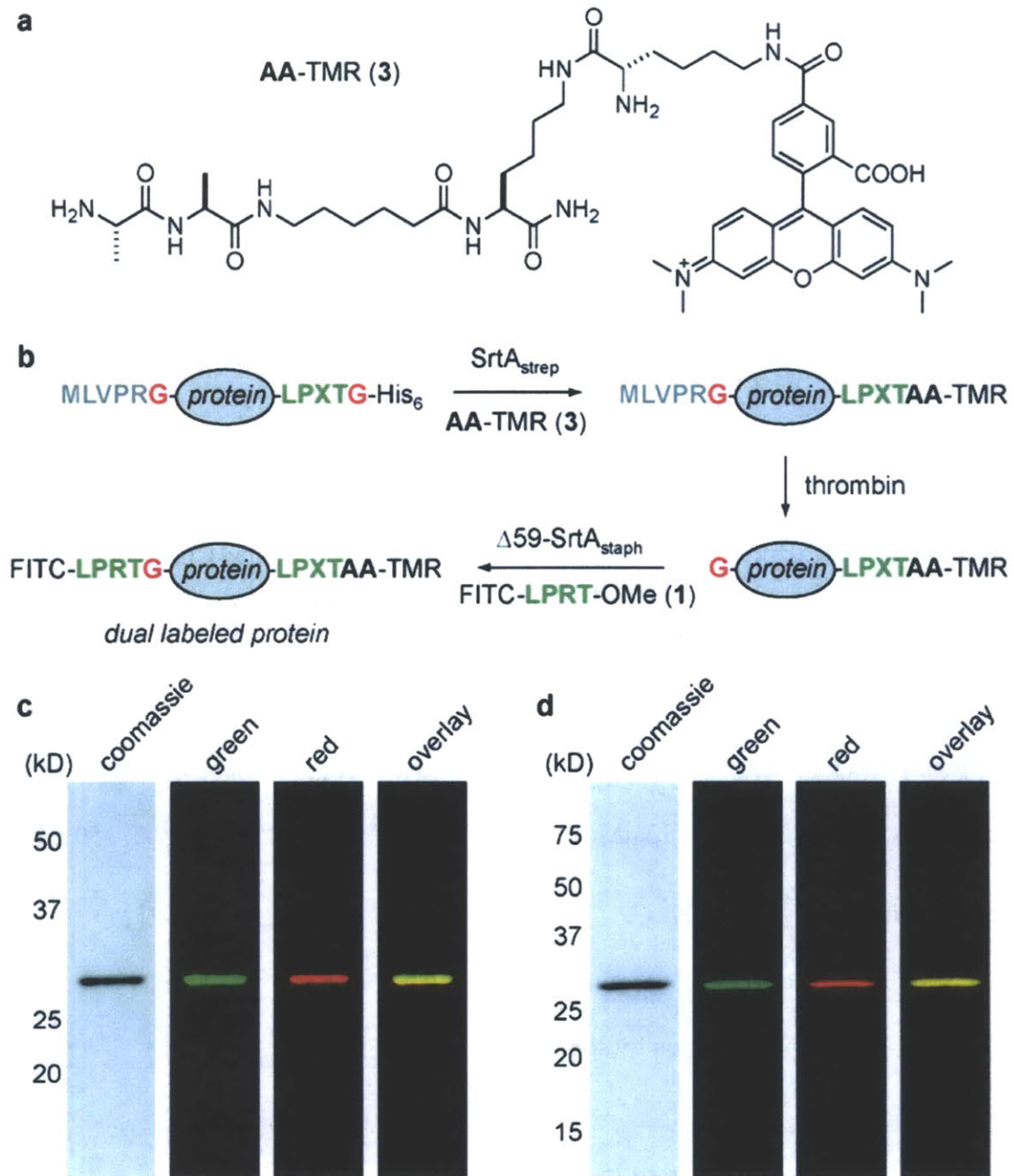


Figure 3.1



*Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities*

Figure 3.2



## **Supplementary Figure Legends**

### **Supplementary Figure 3.1. Synthesis of FITC-LPRT-OMe**

(a) Synthesis of FITC-LPRT-OMe (**1**) and biotin-LPRT-OMe (**2**).

(b) RP-HPLC chromatogram (280 nm) for purified **1** and ESI-MS characterization. (c) RP-HPLC chromatogram (210 nm) for purified **2** and ESI-MS characterization.

### **Supplementary Figure 3.2. Reconstructed ESI-MS spectra for N-terminal labeling reactions**

Reconstructed ESI-MS spectra for N-terminal labeling reactions on (a) G<sub>1</sub>-CtxB, (b) G<sub>3</sub>-CtxB, (c) G<sub>5</sub>-CtxB, and (d) AG<sub>4</sub>-CtxB substrates. Conditions: 50 μM CtxB, 50 μM SrtA<sub>staph</sub>, 500 μM **1**, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2 h at 37 °C.

### **Supplementary Figure 3.3. Site-specific N-terminal biotinylation of G<sub>5</sub>-CtxB.**

Conditions: 33 μM G<sub>5</sub>-CtxB, 50 μM SrtA<sub>staph</sub>, 500 μM **2**, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2 h at 37 °C.

(a) ESI-MS characterization of biotinylation.

(b) Verification of biotinylation by streptavidin immunoblot.

### **Supplementary Figure 3.4. Additional substrates for N-terminal labeling.**

(a) ESI-MS spectra for G<sub>5</sub>-eGFP labeling. Conditions: 50 μM G<sub>5</sub>-eGFP, 50 μM SrtA<sub>staph</sub>, 500 μM **1**, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2h at 37 °C. (b) ESI-MS spectra for G<sub>1</sub>-UCHL3 labeling. Conditions: 50 μM G<sub>1</sub>-UCHL3, 50 μM SrtA<sub>staph</sub>, 500

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

$\mu\text{M}$  **1**, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2 h at 37 °C.

**Supplementary Figure 3.5. Synthesis of AA-TMR**

(a) Synthesis of AA-TMR (**3**).

(b) RP-HPLC chromatogram (280 nm) for purified **3** and ESI-MS characterization.

**Supplementary Figure 3.6. Site-specific labeling at the N and C termini of eGFP and UCHL3.**

(a) ESI-MS spectra for all intermediates generated during the double labeling procedure.

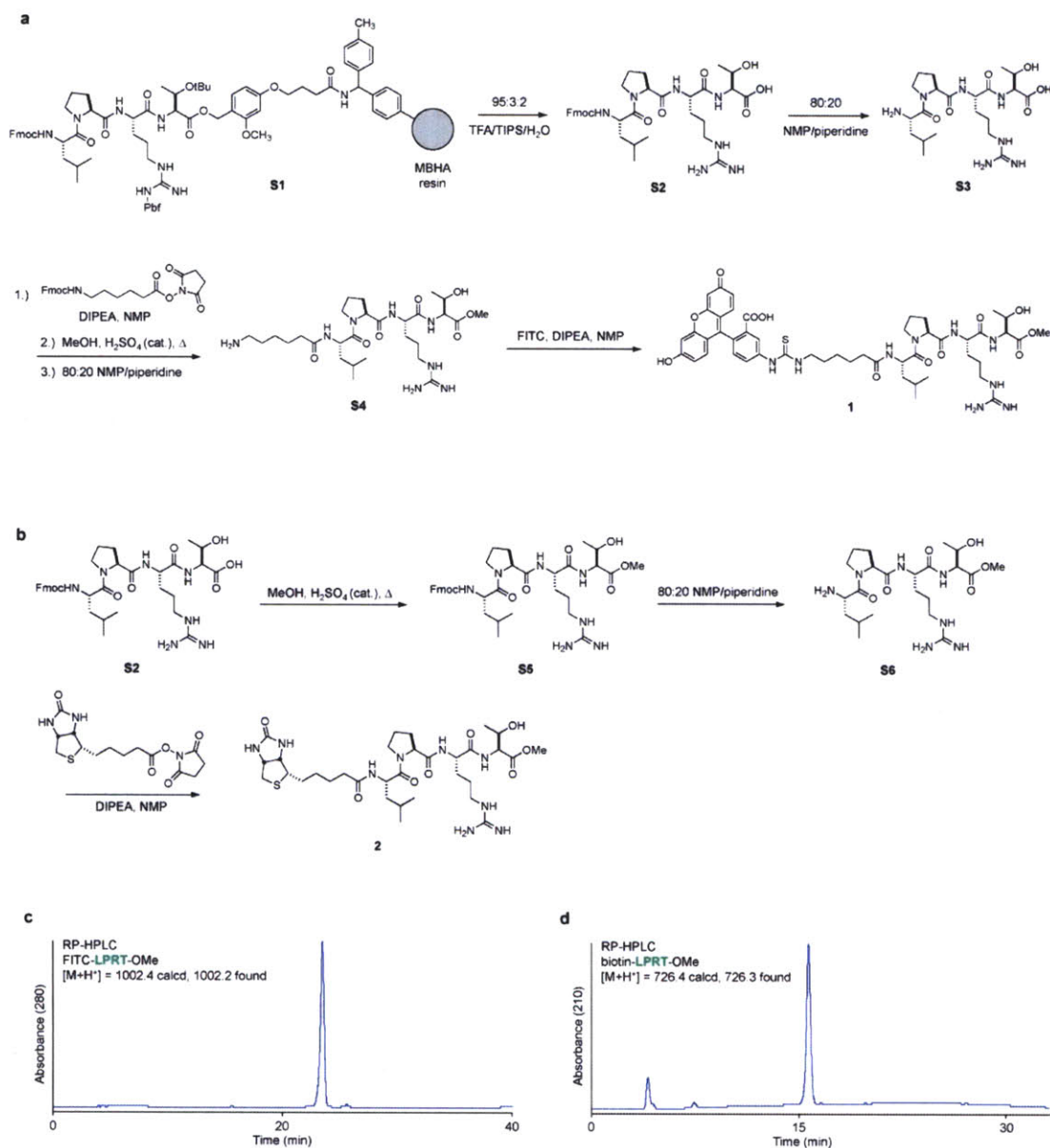
From top to bottom, this includes the eGFP starting material ( $m/z = 31080$  Da), the intermediate formed after C-terminal modification with **3** mediated by SrtA<sub>strep</sub> ( $m/z = 29470$  Da), the product of thrombin cleavage ( $m/z = 28855$  Da), crude dual labeled eGFP ( $m/z = 29726$  Da), and dual labeled eGFP after anion exchange chromatography ( $m/z = 29725$  Da).

(b) ESI-MS spectra for all intermediates generated during double labeling of UCHL3.

From top to bottom, this includes the UCHL3 starting material ( $m/z = 29252$  Da), the intermediate formed after C-terminal modification with **3** mediated by SrtA<sub>strep</sub> ( $m/z = 29050$  Da), the product of thrombin cleavage ( $m/z = 28458$  Da), crude dual labeled UCHL3 ( $m/z = 29412$  Da), and dual labeled UCHL3 after anion exchange chromatography ( $m/z = 29412$  Da). The MTSET reagent used to quench SrtA<sub>strep</sub> also modifies the active site cysteine of UCHL3 creating an extra +118 Da signal in the mass spectrum. This modification is easily removed from the final product by brief treatment

with DTT.

### Supplementary Figure 3.1

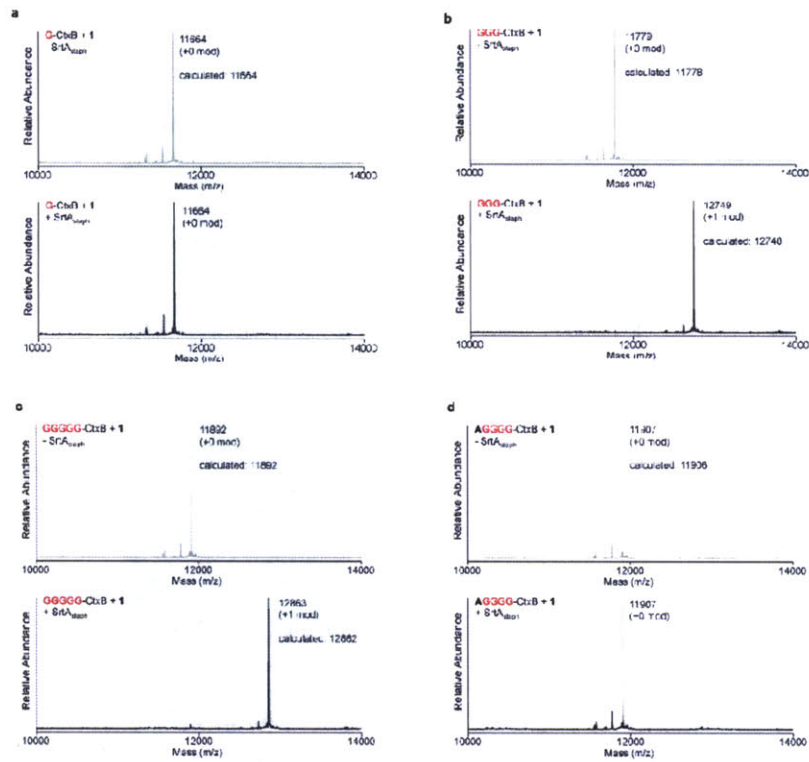


**Figure S1.** (a) Synthesis of FITC-LPRT-OMe (**1**) and biotin-LPRT-OMe (**2**). (b) RP-HPLC chromatogram (280 nm) for purified **1** and ESI-MS characterization. (c) RP-HPLC chromatogram (210 nm) for purified **2** and ESI-MS characterization.



Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

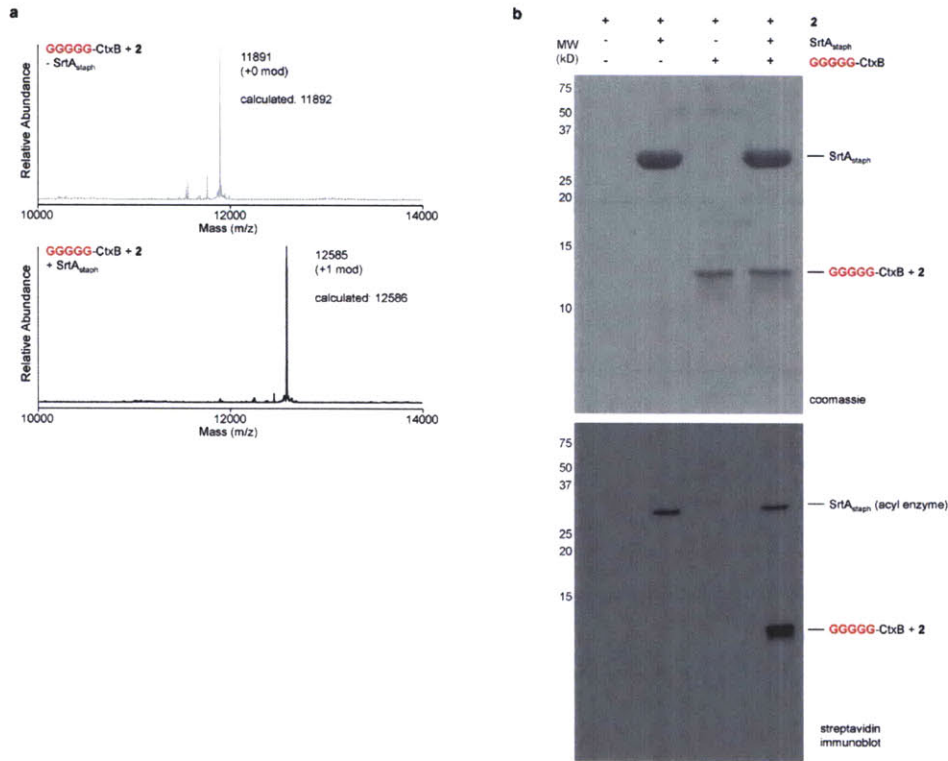
**Supplementary Figure 3.2**



**Figure S2.** Reconstructed ESI-MS spectra for N-terminal labeling reactions on (a) G<sub>1</sub>-CtxB, (b) G<sub>3</sub>-CtxB, (c) G<sub>5</sub>-CtxB, and (d) AG<sub>4</sub>-CtxB substrates. Conditions: 50 μM CtxB, 50 μM SrtA<sub>saph</sub>, 500 μM I, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2 h at 37 °C.

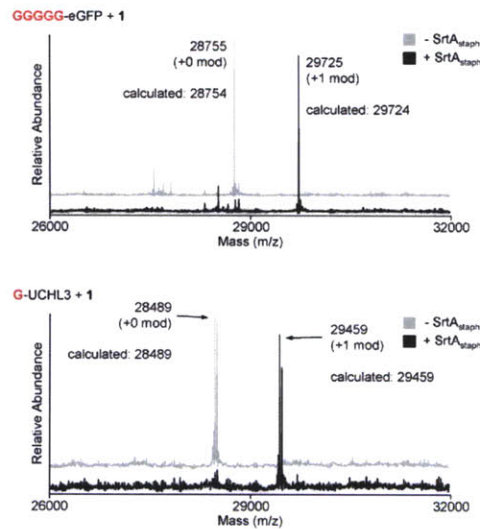
*Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities*

**Supplementary Figure 3.3**



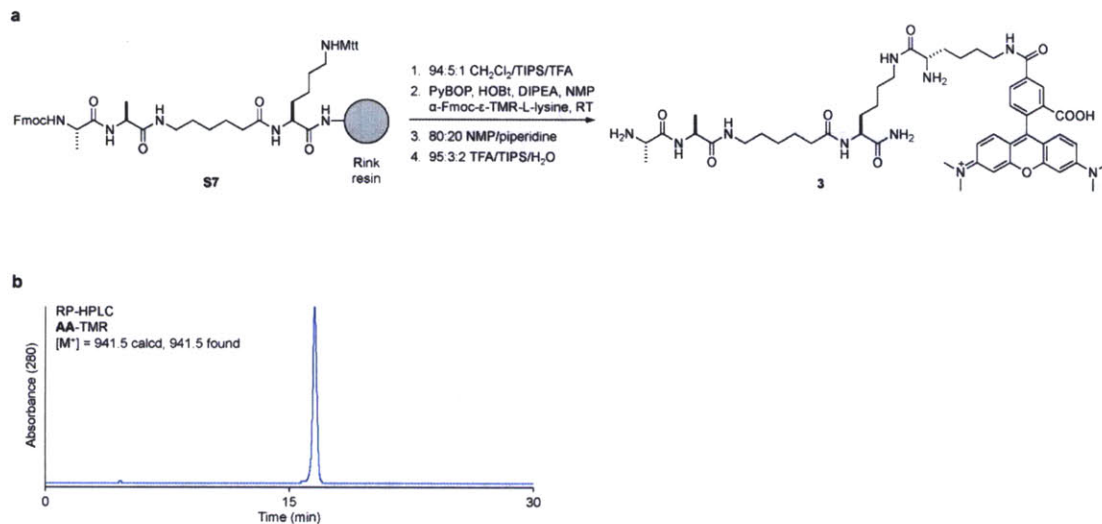
**Figure S3.** Site-specific N-terminal biotinylation of G<sub>5</sub>-CtxB. Conditions: 33 μM G<sub>5</sub>-CtxB, 50 μM SrtA<sub>staph</sub>, 500 μM **2**, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2 h at 37 °C. (a) ESI-MS characterization of biotinylation. (b) Verification of biotinylation by streptavidin immunoblot.

**Supplementary Figure 3.4**



**Figure S4.** Additional substrates for N-terminal labeling. (a) ESI-MS spectra for G<sub>5</sub>-eGFP labeling. Conditions: 50 μM G<sub>5</sub>-eGFP, 50 μM SrtA<sub>staph</sub>, 500 μM **1**, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2 h at 37 °C. (b) ESI-MS spectra for G<sub>1</sub>-UCHL3 labeling. Conditions: 50 μM G<sub>1</sub>-UCHL3, 50 μM SrtA<sub>staph</sub>, 500 μM **1**, 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 2 h at 37 °C.

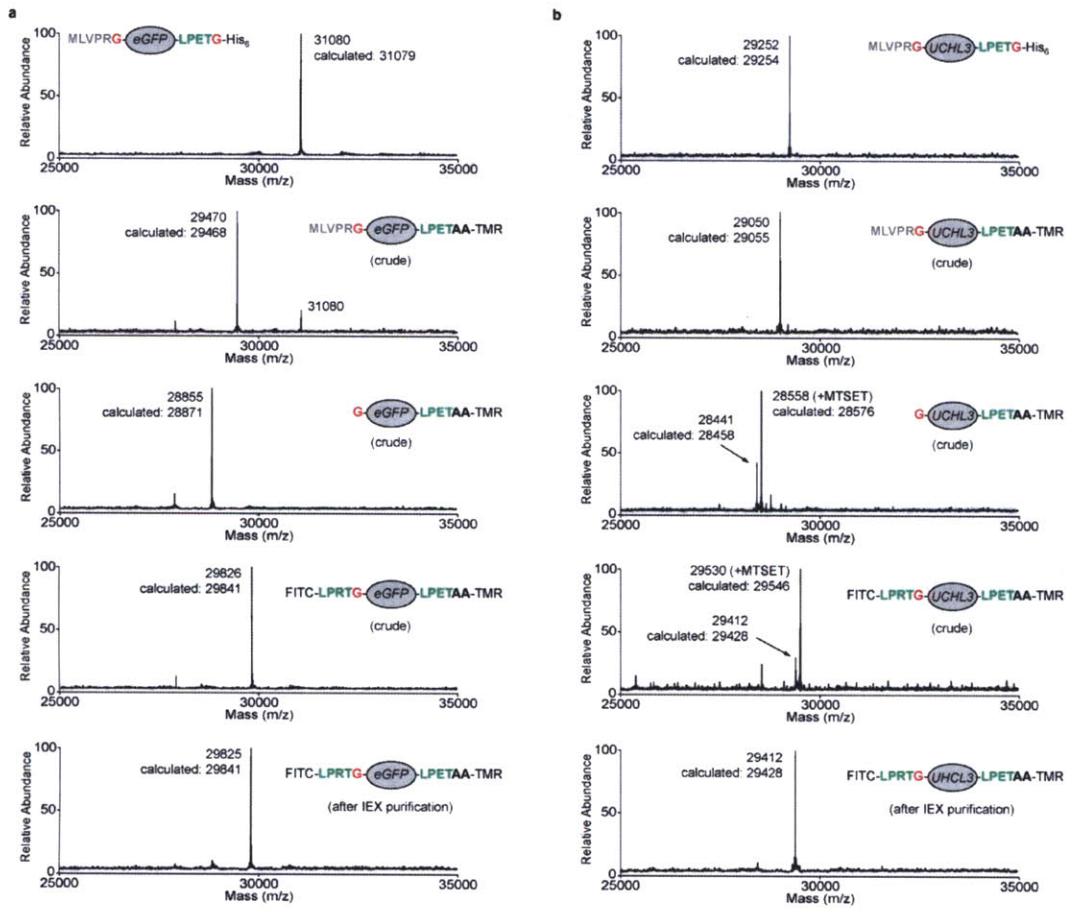
**Supplementary Figure 3.5**



**Figure S5.** (a) Synthesis of AA-TMR (**3**). (b) RP-HPLC chromatogram (280 nm) for purified **3** and ESI-MS characterization.

*Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities*

**Supplementary Figure 3.6**



Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

**Supplemental Text**

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. 4-Methylbenzylhydramine (MBHA) resin HL hydrochloride salt (100-200 mesh, 1.1 mmol/g), 4-(4-hydroxymethyl-3-methoxyphenoxy)-butyric acid (HMPB), Fmoc-Leu-OH, FmocPro-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Ala-OH, and Fmoc- $\epsilon$ -Ahx-OH were obtained from EMB Biosciences/Novabiochem. Rink amide resin (100-200 mesh, 0.7 mmol/g) was obtained from Advanced Chemtech. FITC isomer I was purchased from Sigma (F7250).  $\alpha$ -Fmoc- $\epsilon$ -TMR-Llysine was purchased from Invitrogen. Water used in biological procedures or as a reaction solvent was purified using a MilliQ purification system (Millipore).

DriSolv<sup>®</sup> anhydrous CH<sub>2</sub>Cl<sub>2</sub> and DriSolv<sup>®</sup> anhydrous MeOH were purchased from EMD Chemicals. Redistilled, anhydrous *N,N'*diisopropylethylamine (DIPEA) was obtained from Sigma-Aldrich.

**Mass Spectrometry.** LC-ESI-MS analysis was performed using a Micromass LCT mass spectrometer (Micromass<sup>®</sup> MS Technologies, USA) and a Paradigm MG4 HPLC system equipped with a HTC PAL autosampler (Michrom BioResources, USA) and a Waters Symmetry 5  $\mu$ m C8 column (2.1 x 50 mm, MeCN:H<sub>2</sub>O (0.1% formic acid) gradient mobile phase, 150  $\mu$ L/min).

**HPLC/FPLC.** HPLC purifications were achieved using an Agilent 1100 Series HPLC system equipped with a Waters Delta Pak 15  $\mu$ m, 100 Å C18 column (7.8 x 300 mm, MeCN:H<sub>2</sub>O gradient mobile phase, 3 mL/min) as indicated below. Size exclusion and

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

cation exchange chromatography were performed on a Pharmacia AKTA Purifier system equipped with a HiLoad 16/60 Superdex 75 column (Amersham) or a Mono Q 5/50 GL column (Amersham), respectively.

**UV-vis Spectroscopy.** UV-vis spectroscopy was performed on a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, USA). **In-gel Fluorescence.** Fluorescent gel images were obtained using a Typhoon 9200 Variable Mode Imager (GE Healthcare).

**II. Synthesis and Characterization of FITC-LPRT-OMe (1), biotin-LPRT-OMe (2) and AA-TMR (3)**

**Fmoc-LPRT-OH (S2).** MBHA resin (1.05 g, 1.15 mmol) was first washed/swollen with 25 mL of NMP (3x, 3-5 min per wash). The resin was then treated with a solution of HMPB (690 mg, 2.87 mmol), PyBOP (1.49 g, 2.86 mmol), HOBt (387 mg, 2.86 mmol), and DIPEA (1.48 mL, 8.59 mmol) in

11.5 mL of NMP and incubated for 14 h at RT with gentle agitation on a wrist action shaker. The resin was then washed with 25 mL of NMP (3x, 3-5 min per wash) followed by 25 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 10 min per wash). The resin was then dried and 370 mg of the dry resin was transferred to a new solid phase reaction vessel. The resin was washed with CH<sub>2</sub>Cl<sub>2</sub> and then treated with a solution of Fmoc-Thr(tBu)OH (477 mg, 1.20 mmol), DIC (186  $\mu$ L, 1.20 mmol), and DMAP (14 mg, 0.11 mmol) in 5 mL of anhydrous CH<sub>2</sub>Cl<sub>2</sub>.

The resin was incubated for 16 h at RT and then washed with 20 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash). The coupling of Fmoc-Thr(tBu)-OH was then repeated to achieve

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

maximum resin loading. The resin was then washed with 20 mL of NMP (3x, 3-5 min per wash). Deprotection was achieved with 80:20 NMP/piperidine (20 mL) for 20 min at RT followed by washing with 20 mL of NMP (3x, 3-5 min per wash). The remaining amino acid building blocks (R,P,L) were then coupled as follows: Fmoc-protected amino acid (1.50 mmol, 5.0 equivalents relative to estimated resin loading), PyBOP (781 mg, 1.50 mmol), HOBt (203 mg, 1.50 mmol), and DIPEA (775  $\mu$ L, 4.50 mmol) were dissolved in NMP to a final volume of 7.00 mL. This solution was mixed until all reagents had dissolved, and then added to the deprotected resin. Couplings were incubated for 12-24 h at RT. The resin was then washed with ~30 mL of NMP (3x, 3-5 min per wash). The extent of coupling was assessed by Kaiser test. In the event that the coupling was incomplete, the above procedure was repeated. Fmoc removal was then achieved by exposing the resin to 20 mL of 80:20 NMP/piperidine for 20 min at RT, followed by additional washing with ~30 mL of NMP (3x, 3-5 min per wash). Repeated cycles of amino acid coupling and Fmoc deprotection were then repeated to complete the synthesis of resin-bound intermediate **S1**. Resin **S1** was cleaved by treatment with ~5 mL of 95:3:2 TFA/TIPS/H<sub>2</sub>O (5x, ~30 min each) and the combined cleavage solutions were concentrated. Crude **S2** was precipitated from cold diethyl ether and dried (184 mg, 87% yield based on estimated resin loading). The identity of **S2** was confirmed by ESI-MS ( $[M+H]^+$  = 708.4 calcd, 708.4 obsd). **S2** was used without further purification.

**H2N-LPRT-OH (S3)**. Peptide **S2** (28 mg, 40  $\mu$ mol) was dissolved in 160  $\mu$ L of NMP and treated with 30  $\mu$ L of piperidine. The reaction was incubated at room temperature for 20 min. Crude **S3** was precipitated from cold diethyl ether followed by washing of the

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

resulting solid with three additional portions of diethyl ether. The solid was then dried and used without further purification (8 mg, 41%). The identity of **S3** was confirmed by ESI-MS ( $[M+H]^+$  = 486.3 calcd, 486.1 obsd).

**H2N-Ahx-LPRT-OMe (S4).** Intermediate **S3** (8.0 mg, 16  $\mu$ mol) was combined with Fmocaminohexanoic acid NHS ester (Fmoc-Ahx-NHS)<sup>1</sup> (8.1 mg, 18  $\mu$ mol), and DIPEA (15  $\mu$ L, 86  $\mu$ mol) in 172  $\mu$ L of NMP. The reaction was incubated at room temperature for 19 h. The crude product was then precipitated from cold diethyl ether and the resulting solid washed with two additional portions of diethyl ether. The solid was then dissolved in 750  $\mu$ L of anhydrous MeOH and treated with 5  $\mu$ L of concentrated H<sub>2</sub>SO<sub>4</sub>. The reaction was incubated for 4 h at 50 °C. Formation of the desired methyl ester was verified by ESI-MS. The reaction was then diluted with 20 mL of H<sub>2</sub>O and passed over a Waters Sep-Pak<sup>®</sup> Plus C18 Cartridge (cartridge was pre-equilibrated with 20 mL of 1:1 MeCN/H<sub>2</sub>O followed by 20 mL of H<sub>2</sub>O). The cartridge was washed with 20 mL of H<sub>2</sub>O and the crude methyl ester was eluted with 10 mL of 1:1 MeCN/H<sub>2</sub>O. This material was then concentrated and treated with 100  $\mu$ L of 80:20 NMP/piperidine for 20 min at RT. Crude **S4** was then precipitated from cold diethyl ether, washed with two additional portions of diethyl ether, and dried under vacuum. **S4** was used without further purification (7 mg, 71% from **S3**). The identity of **S4** was confirmed by ESI-MS ( $[M+H]^+$  = 613.4 calcd, 613.3 obsd).

**FITC-LPRT-OMe (1).** Intermediate **S4** (7.0 mg, 11  $\mu$ mol) was combined with FITC isomer I (4.9 mg, 13  $\mu$ mol), and DIPEA (15  $\mu$ L, 86  $\mu$ mol) in 120  $\mu$ L of NMP. The



Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

reaction was incubated at room temperature for 2 h. Crude **1** was then precipitated from cold diethyl ether and the resulting solid washed with two additional portions of diethyl ether. The material was then purified by RP-HPLC [Waters C18 column, MeCN:H<sub>2</sub>O gradient mobile phase, 3 mL/min, 5% MeCN (0-2 min), 5% MeCN → 40% MeCN (2-25 min)] to yield **1** (3.8 mg, 35%). The identity and purity of **1** was confirmed by RPHPLC and ESI-MS (**Figure S1b**). A 10 mM stock solution of **1** in DMSO was used in all transpeptidation experiments.

**Fmoc-LPRT-OMe (S5)**. Peptide **S2** (35 mg, 49 μmol) was dissolved in 750 μL of anhydrous MeOH and treated with 5 μL of concentrated H<sub>2</sub>SO<sub>4</sub>. The reaction was incubated for 3 h at 55 °C followed by an additional 15 h at RT. The reaction was then diluted with 20 mL of H<sub>2</sub>O and passed over a Waters Sep-Pak<sup>®</sup> Plus C18 Cartridge (cartridge was pre-equilibrated with 20 mL of 1:1 MeCN/H<sub>2</sub>O followed by 20 mL of H<sub>2</sub>O). The cartridge was washed with 20 mL of H<sub>2</sub>O and then crude **S5** was eluted with 10 mL of 1:1 MeCN/H<sub>2</sub>O. **S5** was then concentrated and used without further purification (24 mg, 68%). The identity of **S5** was confirmed by ESI-MS ([M+H]<sup>+</sup> = 722.4 calcd, 722.5 obsd).

**H<sub>2</sub>N-LPRT-OMe (S6)**. Methyl ester **S5** (23 mg, 32 μmol) was dissolved in 125 μL of NMP and treated with 30 μL of piperidine. The reaction was incubated at room temperature for 20 min. Crude **S6** was precipitated from cold diethyl ether followed by washing of the resulting solid with three additional portions of diethyl ether. The solid was then dried and used without further purification (16 mg, quantitative). The identity

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

of **S6** was confirmed by ESI-MS ( $[M+H]^+$  = 500.3 calcd, 500.2 obsd).

**Biotin-LPRT-OMe (2).** Intermediate **S6** (5.0 mg, 10  $\mu$ mol) was combined with biotin-NHS<sup>2</sup> (3.8 mg, 11  $\mu$ mol), and DIPEA (10  $\mu$ L, 58  $\mu$ mol) in 105  $\mu$ L of NMP. The reaction was incubated at room temperature for 18 h. The material was then purified by RP-HPLC [Waters C18 column, MeCN:H<sub>2</sub>O gradient mobile phase, 3 mL/min, 5% MeCN (0-2 min), 5% MeCN  $\rightarrow$  60% MeCN (2-30 min)] to yield **2** (3 mg, 41%). The identity and purity of **2** was confirmed by RP-HPLC and ESI-MS (**Supplementary Figure 3.1d**). A 10 mM stock solution of **2** in DMSO was used in all transpeptidation experiments.

**AA-TMR (3).** Resin bound intermediate **S7** was synthesized on Rink amide resin using standard Fmoc synthesis. Couplings were performed using 5 equivalents (relative to estimated resin loading) of the suitably protected Fmoc amino building block, 5 equivalents of PyBOP, 5 equivalents of HOBt, and 15 equivalents of DIPEA in NMP (final concentration of Fmoc amino acid was ~170 mM). Couplings were run for 5-72 h at room temperature with gentle agitation on a wrist action shaker. Fmoc deprotection was achieved with 80:20 NMP/piperidine for 20 min at room temperature. The resin was washed with NMP between each transformation. Following completion of **S5**, the resin was washed with NMP (4x, 3-5 min per wash) and CH<sub>2</sub>Cl<sub>2</sub> (5x, 3-5 min per wash). The resin was then dried. Dry **S5** resin (43 mg, 21  $\mu$ mol, 0.5 mmol/g estimated resin loading) was then transferred to a 3.0 mL fritted polypropylene syringe equipped with a hypodermic needle. The 4-methyltrityl (Mtt) protecting group was removed by treatment with 2.5 mL of 94:5:1 CH<sub>2</sub>Cl<sub>2</sub>/TIPS/TFA at

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

room temperature (5x, 5 min each) followed by washing with 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash) and 2.5 mL of NMP (3x, 3-5 min per wash). The resin was then treated with a solution of  $\alpha$ -Fmoc- $\epsilon$ -TMR-L-lysine (25 mg, 32  $\mu$ mol), PyBOP (17 mg, 33  $\mu$ mol), HOBt (9 mg, 70  $\mu$ mol), and DIPEA (16.5  $\mu$ L, 95.8  $\mu$ mol) in 1.0 mL of NMP. The reaction was incubated at RT for 20 h followed by washing with 2.5 mL of NMP (3x, 3-5 min per wash). Fmoc removal was achieved by treatment with 2.5 mL of 80:20 NMP/piperidine for 20 min at RT followed by washing with 2.5 mL of NMP (3x, 3-5 min per wash) and 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x, 3-5 min per wash). The peptide was cleaved from the resin with 2.5 mL of 95:3:2 TFA/TIPS/H<sub>2</sub>O (5x, ~15 min each) and the combined cleavage solutions were concentrated *in vacuo*. Crude **3** was precipitated from cold diethyl ether, and then purified by RP-HPLC [Waters C18 column, MeCN:H<sub>2</sub>O gradient mobile phase containing 0.1% TFA, 3 mL/min, 5% MeCN (0-2 min), 5% MeCN  $\rightarrow$  60% MeCN (2-30 min)]. The identity and purity of **3** (8 mg, 40%) was confirmed by RP-HPLC and ESI-MS (**Supplementary Figure 5.b**). A 100 mM stock solution of **3** in water was used for transpeptidation experiments.

### **III. Protein Cloning and Expression**

**SrtA<sub>strep</sub>**. The expression plasmid for SrtA<sub>strep</sub> (residues 82-249) including an N-terminal His<sub>6</sub> tag has been described.<sup>3</sup> The construct was transformed into *E. coli* BL-21. Cells were grown in 2 L of sterile LB containing kanamycin (30  $\mu$ g/mL) to an optical density of ~0.7 at 600 nm. Cells were induced with IPTG (1 mM) for 3 h at 37 °C. Cells were harvested by centrifugation and the pellet was stored overnight at -20 °C. The pellet was thawed and resuspended in 70 mL of 50 mM Tris pH 8.0, 150 mM NaCl, 20 mM

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

imidazole and 10% glycerol. Cells were then treated with 300  $\mu$ L of DNase I (10 mg/mL in PBS), 500  $\mu$ L of lysozyme (50 mg/mL in PBS), and 10  $\mu$ L of MgCl<sub>2</sub> (1 M in PBS).

The lysis reaction was incubated for 1 h at 4 °C. The cells were then sonicated and centrifuged to remove insoluble material. The clarified lysate was then applied to a Ni-NTA column consisting of 5.0 mL of commercial Ni-NTA slurry (Qiagen) equilibrated with 50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole, and 10% glycerol. The column was washed with 80 mL of 50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole, and 10% glycerol. Protein was eluted with five 5 mL portions of 50 mM Tris pH 8.0, 150 mM NaCl, 300 mM imidazole, and 10% glycerol. Fractions containing SrtA<sub>strep</sub> were pooled and further purified by size exclusion chromatography on a HiLoad 16/60 Superdex 75 column (Amersham), eluting with 20 mM Tris pH 8.0, 150 mM NaCl at a flow rate of 1 mL/min. Fractions containing SrtA<sub>strep</sub> were pooled and subjected to a second round of Ni-affinity chromatography. Purified SrtA<sub>strep</sub> was then dialyzed against 50 mM Tris pH 8.0, 150 mM NaCl, and 10% glycerol. These solutions were stored at -80 °C until further use. Protein concentration was estimated by Bradford assay.

**SrtA<sub>staph</sub>.** Recombinant SrtA<sub>staph</sub> (residues 26-206) containing an N-terminal His<sub>6</sub> tag was produced in *E. coli* as previously described.<sup>4</sup> SrtA<sub>staph</sub> does not contain an N-terminal glycine residue (retains initiator methionine). Purified SrtA<sub>staph</sub> was stored in 10% (w/v) glycerol, 50 mM Tris pH 8.0, 150 NaCl at -80 °C until further use. Protein concentration was estimated by Bradford assay.

**$\Delta$ 59-SrtA<sub>staph</sub>.** Recombinant  $\Delta$ 59-SrtA<sub>staph</sub> (residues 60-206) containing an N-terminal His<sub>6</sub> tag was cloned into pET28a+.  $\Delta$ 59-SrtA<sub>staph</sub> does not contain an N-terminal glycine

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

residue (retains initiator methionine). Expression of  $\Delta 59$ -SrtA<sub>staph</sub> was achieved following the protocol described above for SrtA<sub>strep</sub>. Purification by size exclusion chromatography was not necessary because  $\Delta 59$ -SrtA<sub>staph</sub> was sufficiently pure following Ni-affinity chromatography.

**CtxB.** The template for construction of G1-CtxB, G3-CtxB, G5-CtxB, and AG4-CtxB consisted of the B-subunit of cholera toxin fused at its N terminus to the signal peptide sequence of *E. coli* heat labile enterotoxin LTIIb<sup>5</sup>. This targets the expressed protein to the periplasm where the signal peptide is removed. Glycine and/or alanine residues were inserted between the signal sequence and CtxB via Quickchange® II Site-Directed Mutagenesis (Stratagene). Plasmids were transformed into *E. coli* BL

21. Cells were grown in 1 L of sterile LB containing chloroamphenicol (34  $\mu\text{g}/\text{mL}$ ) to an optical density of  $\sim 0.5$ - $1.0$  at 600 nm. Cells were induced with arabinose (0.25% w/v) for 3 h at 37 °C. Cells were harvested by centrifugation and the pellet was stored overnight at -20 °C. The pellet was thawed and resuspended in 30 mL of 50 mM Tris pH 8.0 and 300 mM NaCl. This suspension was then treated with 3 mL of polymixin B solution (5 mg/mL freshly made in water). This mixture was gently stirred at room temperature for 1 h and then centrifuged. The clarified lysate was treated with 2.5 mL of Ni-NTA slurry (Qiagen). CtxB has a naturally affinity for Ni-NTA although it does not possess a His<sub>6</sub> tag. The Ni-NTA mixture was incubated at 4 °C for 1 h and then poured into a fritted plastic column (Bio-Rad). The resin was washed with 40 mL of 50 mM Tris pH 8.0 and 300 mM NaCl. Protein was eluted with two 10 mL portions of 50 mM Tris pH 8.0, 300 mM NaCl, and 300 mM imidazole. Purified CtxB was buffer exchanged into 20 mM Tris

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

pH 8.0 and 150 mM NaCl. Solutions were stored at 4 °C. Protein concentration was estimated by Bradford assay.

**G<sub>5</sub>-eGFP and dual labeling eGFP substrate.** G<sub>5</sub>-eGFP (containing a C-terminal His<sub>6</sub> tag) and the eGFP dual labeling substrate (containing an N-terminal thrombin cleavage site and a C-terminal LPETG motif followed by a His<sub>6</sub> tag) and were prepared in pET28a+ (Novagen) using a Quickchange® II Site-Directed Mutagenesis Kit (Stratagene). The template plasmid used for mutagenesis has been described.<sup>6</sup> Plasmids were then transformed into *E. coli* BL-21. In a typical experiment, cells were grown in sterile LB containing kanamycin (30 µg/mL) to an optical density of ~0.6-0.9 at 600 nm. Cells were induced with IPTG (1 mM) for 3 h at 37 °C. Cells were harvested by centrifugation and the pellet was stored overnight at -20 °C. The pellet was thawed and resuspended in 20 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole and 1% NP-40. The cell suspension was then lysed by French press and centrifuged. The clarified lysate was then applied to a Ni-NTA column consisting of 5.0 mL of commercial Ni-NTA slurry (Qiagen) equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole and 1% NP-40. The column was washed with 40 mL of 20 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole and 1% NP-40, followed by 40 mL of 20 mM Tris pH 8.0, 150 mM NaCl, and 20 mM imidazole. Protein was eluted with 20 mM Tris pH 8.0, 150 mM NaCl, and 300 mM imidazole until the characteristic green color was fully removed from the column. This material was concentrated and further purified by size exclusion chromatography on a HiLoad 16/60 Superdex 75 column (Amersham), eluting with 20 mM Tris pH 8.0, 150 mM NaCl at a flow rate of 1 mL/min. Fractions containing eGFP

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

were pooled and subjected to a second round of Ni-affinity chromatography. Purified eGFP was then buffer exchanged into 20 mM Tris pH 8.0, 150 mM NaCl using a PD-10 Sephadex<sup>TM</sup> column (GE Healthcare), concentrated, and treated with glycerol (10% v/v final concentration). These solutions were stored at 80 °C until further use. Protein concentration was estimated by UV-vis spectroscopy using the absorbance of eGFP at 488 nm (extinction coefficient 55,900 M<sup>-1</sup> cm<sup>-1</sup>).

**UCHL3 and dual labeling UCHL3 substrate.** UCHL3 containing a single N-terminal glycine residue was produced in *E. coli* as described previously.<sup>8</sup> This construct (in pET28a+, Novagen) was then used to prepare the dual labeling UCHL3 substrate. Synthetic 5'-phosphorylated oligonucleotide duplexes containing appropriate sticky ends were designed to achieve insertion of an N-terminal thrombin site and a C-terminal LPETG sequence separated from UCHL3 by a GGGGSGGGGS spacer in two sequential cloning steps. Duplexes were annealed before ligation into the parent vector. The C-terminal insertion was performed first using the PstI and XhoI restriction sites. The result of using the XhoI site was the addition of a His6-tag after the LPETG sequence. The N-terminal insertion was then achieved using the XbaI and NdeI restriction sites. This plasmid was then transformed into *E. coli* BL21. Cells were grown in sterile LB containing kanamycin (30 µg/mL) to an optical density of ~0.6-0.9 at 600 nm. Cells were induced with IPTG (1 mM) for 3 h at 37 °C. Bacteria were then harvested by centrifugation and the pellet was stored overnight at -20 °C. The pellet was thawed and resuspended in 20 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole and 1% NP-40.

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

The cell suspension was then lysed by French press and centrifuged. The clarified lysate was then applied to a Ni-NTA column consisting of 5.0 mL of commercial Ni-NTA slurry (Qiagen) equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole and 1% NP-40. The column was washed with 40 mL of 20 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole and 1% NP-40, followed by 40 mL of 20 mM Tris pH 8.0, 150 mM NaCl, and 20 mM imidazole. Protein was eluted with 20 mM Tris pH 8.0, 150 mM NaCl, and 300 mM imidazole. This material was then purified by anion-exchange chromatography on a Mono Q 5/50 GL column (Amersham) [Buffer A (50 mM Tris pH 7.5, 5 mM DTT, 0.5 mM EDTA), Buffer B (50 mM Tris pH 7.5, 5 mM DTT, 0.5 mM EDTA, 500 mM NaCl), 1.5 mL/min, gradient: 100% Buffer A (0-15 mL), 0% Buffer B → 50% Buffer B (15-45 mL), 50% Buffer B (45-50 mL)]. Fractions containing UCHL3 were pooled and further purified by size exclusion chromatography on a HiLoad 16/60 Superdex 75 column (Amersham), eluting with 20 mM Tris pH 8.0, 150 mM NaCl at a flow rate of 1 mL/min. Fractions containing UCHL3 were pooled and subjected to a final purification step by anion-exchange chromatography on a Mono Q 5/50 GL column (Amersham) [Buffer A (50 mM phosphate pH 6.0), Buffer B (50 mM phosphate pH 6.0, 500 mM NaCl), 1.5 mL/min, gradient: 100% Buffer A (0-15 mL), 0% Buffer B → 50% Buffer B (15-45 mL), 50% Buffer B (45-50 mL)]. The dual labeling UCHL3 substrate was buffer exchanged into 20 mM Tris pH 8.0, 150 NaCl and protein concentration was estimated by Bradford assay.

#### **IV. Sortase-Mediated Labeling of Protein Substrates**

**N-terminal labeling.** N-terminal transpeptidation reactions were performed by



Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

combining the necessary proteins/reagents at the specified concentrations in the presence of SrtA<sub>staph</sub> or  $\Delta 59$ -SrtA<sub>staph</sub> in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) and incubating at 37 °C for the times indicated. Reactions were either diluted with 2x reducing Laemmli sample buffer for SDS-PAGE analysis or diluted with water (~50 fold) for ESI-MS analysis. Gels were visualized by staining with coomassie blue. Fluorescence was visualized on a Typhoon 9200 Imager (GE Healthcare). For detection of biotinylation, proteins were separated by SDS-PAGE and transferred to a nitrocellulose membrane. The membrane was then probed with a streptavidin-horseradish peroxidase conjugate (GE Healthcare) and visualized by chemiluminescence. ESI-MS was performed on a Micromass LCT mass spectrometer (Micromass® MS Technologies, USA) and a Paradigm MG4 HPLC system equipped with a HTC PAL autosampler (Michrom BioResources, USA) and a Waters Symmetry 5  $\mu$ m C8 column (2.1 x 50 mm, MeCN:H<sub>2</sub>O (0.1% formic acid) gradient mobile phase, 150  $\mu$ L/min).

**Dual Labeling of eGFP**

Immediately prior to starting the dual labeling sequence, the eGFP stock was thawed and again purified by affinity chromatography over commercial Ni-NTA resin. After binding eGFP to the resin, the column was washed with 20 mM Tris pH 8.0, 150 mM NaCl, and 20 mM imidazole. The protein was eluted with 20 mM Tris pH 8.0, 150 mM NaCl, and 300 mM imidazole. This material was buffer exchanged into 20 mM Tris pH 8.0, 150 mM NaCl using a NAP<sup>TM</sup> 5 Sephadex<sup>TM</sup> column (GE Healthcare) and concentrated. The concentration was estimated to be 84  $\mu$ M by UV-vis spectroscopy using eGFP

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

absorbance at 488 nm (extinction coefficient  $55,900 \text{ M}^{-1} \text{ cm}^{-1}$ ).

*C-terminal modification of eGFP with 3 and SrtA<sub>strep</sub>.* 400  $\mu\text{L}$  of the freshly purified eGFP solution was then treated with SrtA<sub>strep</sub> (87  $\mu\text{L}$  of a 140  $\mu\text{M}$  stock solution) and **3** (4.9  $\mu\text{L}$  of a 100 mM stock solution) [**Note:** SrtA<sub>strep</sub> does not require  $\text{Ca}^{2+}$  for activity].<sup>3</sup>

The reaction was incubated for 7 h at 37 °C. ESI-MS analysis of the crude reaction mixture revealed excellent conversion to the desired product (Supporting Figure S6a). The reaction was then treated with [2-(trimethylammonium)ethyl] methane thiosulfonate bromide (MTSET) (2.5  $\mu\text{L}$  of a 500 mM solution in 1:1 DMSO/H<sub>2</sub>O) for 10 min at room temperature to quench SrtA<sub>strep</sub>. The entire reaction was then diluted with 5 mL of 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. This solution was then passed over a 1.5 mL column of Ni-NTA that been equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. The column was then washed with 1.5 mL of 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. His<sub>6</sub>-tagged SrtA<sub>strep</sub> was bound by Ni-NTA while the eGFP product (which lost its His<sub>6</sub> tag during the course of transpeptidation) was not retained. The eGFP solution was then concentrated and passed over a PD-10 Sephadex<sup>TM</sup> desalting column (equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl) to remove excess **3**. This material was concentrated to ~800  $\mu\text{L}$  and subjected to thrombin cleavage.

*Thrombin cleavage of eGFP.* All 800  $\mu\text{L}$  of the solution described above was combined with 100  $\mu\text{L}$  of 10x cleavage buffer and 100  $\mu\text{L}$  of thrombin agarose beads (from

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

Thrombin CleanCleave™ Kit, Sigma). This mixture was incubated for 1 h at 37 °C, and then checked by ESI-MS to ensure quantitative cleavage (Supporting Figure S6a). The reaction was then filtered to remove the thrombin beads.

*N-terminal labeling of eGFP with 1 and  $\Delta 59$ -SrtA<sub>staph</sub>.* 389  $\mu$ L of the thrombin cleaved material was combined with **1** (25  $\mu$ L of a 10 mM DMSO solution),  $\Delta 59$ -SrtA<sub>staph</sub> (36  $\mu$ L of a 700  $\mu$ M stock solution), and 10x sortase reaction buffer (50  $\mu$ L of 500 mM Tris pH 8.0, 1.5 M NaCl, 100 mM CaCl<sub>2</sub>). The reaction was incubated for 75 min at 37 °C. ESI-MS of the crude reaction mixture showed clean formation of the dual labeled product as the major reaction product (Supporting Figure S6a). The reaction was then diluted with 5 mL of 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. This solution was passed over a 1.5 mL column of Ni-NTA that been equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole in order to remove His6-tagged  $\Delta 59$ -SrtA<sub>staph</sub>. 2.5 mL of this eluate was then passed over a PD-10 Sephadex™ desalting column (equilibrated with 20 mM Tris pH 8.0). This material was then purified by anion-exchange chromatography on a Mono Q 5/50 GL column (Amersham) [Buffer A (20 mM Tris pH 8.0), Buffer B (20 mM Tris pH 8.0, 1 M NaCl), 1.5 mL/min, gradient: 100% Buffer A (0-15 mL), 0% Buffer B → 50% Buffer B (15-45 mL), 50% Buffer B (45-50 mL)]. Fractions containing dual labeled eGFP were pooled and analyzed by SDS-PAGE and ESI-MS. Coomassie stained gels were imaged using a CanoScan 8600F scanner. Protein purity was estimated from these images using ImageJ 1.42q densitometry software.

### Dual labeling of UCHL3

*C-terminal modification of UCHL3 with 3 and SrtA<sub>strep</sub>.* UCHL3 (350  $\mu$ L of a 65  $\mu$ M stock solution) was treated with SrtA<sub>strep</sub> (76  $\mu$ L of a 140  $\mu$ M stock solution) and **3** (4.3  $\mu$ L of a 100 mM stock solution) [**Note:** SrtA<sub>strep</sub> does not require Ca<sup>2+</sup> for activity<sup>3</sup>]. The reaction was incubated for 15 h at 37 °C. ESI-MS analysis of the crude reaction mixture revealed excellent conversion to the desired product (Supporting Figure S6b). The entire reaction was then diluted with 5 mL of 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. This solution was then treated with [2-(trimethylammonium)ethyl] methane thiosulfonate bromide (MTSET) (5.0  $\mu$ L of a 500 mM solution in 1:1 DMSO/H<sub>2</sub>O) for 10 min at room temperature to quench SrtA<sub>strep</sub>. [**Note:** UCHL3 contains an active site cysteine residue and is therefore modified by MTSET. The resulting modification is disulfide linked, and is easily removed by treatment with DTT following completion of the dual labeling procedure]. The diluted reaction solution was then passed over a 1.5 mL column of Ni-NTA that been equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. The column was then washed with 1.5 mL of 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. The UCHL3 solution was then concentrated and passed over a PD-10 Sephadex<sup>TM</sup> desalting column (equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl) to remove excess **3**. This material was concentrated to 1 mL and subjected to thrombin cleavage.

*Thrombin cleavage of UCHL3.* All 1 mL of the solution described above was combined with 100  $\mu$ L of 10x cleavage buffer and 100  $\mu$ L of thrombin agarose beads (from Thrombin CleanCleave<sup>TM</sup> Kit, Sigma). This mixture was incubated for 1 h at 37 °C, and

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

then checked by ESI-MS to ensure quantitative cleavage (Supporting Figure S6b). The reaction was then filtered to remove the thrombin beads.

*N-terminal labeling of UCHL3 with 1 and  $\Delta 59$ -SrtA<sub>staph</sub>.* 778  $\mu$ L of the thrombin cleaved material was combined with **1** (50  $\mu$ L of a 10 mM DMSO solution),  $\Delta 59$ -SrtA<sub>staph</sub> (72  $\mu$ L of a 700  $\mu$ M stock solution), and 10x sortase reaction buffer (100  $\mu$ L of 500 mM Tris pH 8.0, 1.5 M NaCl, 100 mM CaCl<sub>2</sub>). The reaction was incubated for 60 min at 37 °C. ESI-MS of the crude reaction mixture showed clean formation of the dual labeled product as the major reaction product (Supporting Figure S6b). The reaction was then diluted with 5 mL of 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. This solution was passed over a 1.5 mL column of Ni-NTA that been equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole in order to remove His<sub>6</sub>-tagged  $\Delta 59$ -SrtA<sub>staph</sub>. The column was then washed with 2.0 mL of 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole. The eluate was then desalted using PD-10 Sephadex<sup>TM</sup> columns (equilibrated with 20 mM Tris pH 8.0). This material was then purified by anion-exchange chromatography on a Mono Q 5/50 GL column (Amersham) [Buffer A (20 mM Tris pH 8.0), Buffer B (20 mM Tris pH 8.0, 1 M NaCl), 1.5 mL/min, gradient: 100% Buffer A (0-15 mL), 0% Buffer B → 50% Buffer B (15-45 mL), 50% Buffer B (45-50 mL)]. Fractions containing dual labeled UCHL3 were pooled and analyzed by SDS-PAGE and ESI-MS. Prior to ESI-MS, dual labeled UCHL3 was treated with 10 mM DTT for 10 min at RT to remove the MTSET modification on the active site cysteine residue. Coomassie stained gels were imaged using a CanoScan 8600F scanner. Protein purity was estimated from these images using ImageJ 1.42q densitometry software.

Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

**References**

- (1) Zumbuehl, A.; Jeannerat, D.; Martin, S. E.; Sohrmann, M.; Stano, P.; Vigassy, T.; Clark, D. D.; Hussey, S. L.; Peter, M.; Peterson, B. R.; Pretsch, E.; Walde, P.; Carreira, E. M. *Angew Chem Int Ed Engl* **2004**, *43*, 5181-5.
- (2) Kottani, R.; Valiulin, R. A.; Kutateladze, A. G. *Proc Natl Acad Sci U S A* **2006**, *103*, 13917-21.
- (3) Race, P. R.; Bentley, M. L.; Melvin, J. A.; Crow, A.; Hughes, R. K.; Smith, W. D.; Sessions, R. B.; Kehoe, M. A.; McCafferty, D. G.; Banfield, M. J. *J Biol Chem* **2009**, *284*, 6924-33.
- (4) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12424-9.
- (5) Jobling, M. G.; Palmer, L. M.; Erbe, J. L.; Holmes, R. K. *Plasmid* **1997**, *38*, 158-73.
- (6) Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *J Am Chem Soc* **2008**, *130*, 16338-43.
- (7) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.
- (8) Popp, M. W.; Artavanis-Tsakonas, K.; Ploegh, H. L. *Journal of Biological Chemistry* **2008**, *284*, 3593-3602.

Print Article (Reprinted with permission from Journal of the American Chemical Society):

J|A|C|S  
COMMUNICATIONS

Published on Web 07/17/2009

Site-Specific N- and C-Terminal Labeling of a Single Polypeptide Using Sortases of Different Specificity

John M. Antos, Guo-Liang Chew, Carla P. Guimaraes, Nicholas C. Yoder, Gijsbert M. Grotenbreg, Maximilian Wei-Lin Popp, and Hidde L. Ploegh\*

Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142

Received April 3, 2009; E-mail: ploegh@wi.mit.edu

Methods for site-specific modification of proteins remain in high demand. The transpeptidation reaction catalyzed by sortase A from *Staphylococcus aureus* (SrtA<sub>staph</sub>) allows site-specific derivatization of proteins with virtually any type of functional material.<sup>1</sup> Target proteins are engineered to contain the SrtA<sub>staph</sub> recognition site (LPXTG) near their C terminus, thus allowing a transacylation reaction in which the residues C-terminal to threonine are exchanged for a synthetic oligoglycine peptide (Scheme 1). While the range of applications for this technology has expanded considerably,<sup>1</sup> the ligation chemistry itself has seen relatively few modifications or improvements. Since nearly all Gram-positive bacteria possess sortases, many with reactivity distinct from SrtA<sub>staph</sub>, there is an exciting opportunity to develop complementary strategies for protein engineering using other members of this enzyme family.<sup>2,3</sup> Here we present a strategy for placing discrete labels at both termini in the same polypeptide through the use of multiple sortases. We first describe the ability of SrtA<sub>staph</sub> to append labels at the protein N terminus and then demonstrate how this can be used in conjunction with the activity of sortase A from *Streptococcus pyogenes* (SrtA<sub>strep</sub>) to yield dual-labeled proteins.

Scheme 1. C-terminal Labeling Using SrtA<sub>staph</sub>



With regard to N-terminal labeling mediated by SrtA<sub>staph</sub>, we reasoned that labeled synthetic peptides containing the LPXTG recognition motif, or structural analogues thereof, could generate the requisite acyl-enzyme intermediate necessary for transpeptidation. In combination with protein nucleophiles containing one or more N-terminal glycines, this should result in transfer of the label to the protein N terminus (Figure 1a). We synthesized FITC (1) and biotin (2) derivatives of an LPRT peptide in which the glycine of the normal LPXTG motif was replaced by a methyl ester (Figure 1b and Figure S1 in the Supporting Information). The use of an ester derivative rather than the entire LPXTG motif was motivated by our concern that the glycine residue released after LPXTG cleavage might compete with the protein nucleophile, potentially complicating the desired N-terminal-labeling reaction. In contrast, transacylation with 1 and 2 would generate MeOH, a poor nucleophile for transacylation compared with glycine. It should be noted that concurrently with the work described here, it was demonstrated that labeled LPETGG peptides are viable tools for N-terminal labeling using SrtA<sub>staph</sub>.<sup>4</sup>

With 1 and 2 in hand, we expressed a series of model proteins containing N-terminal glycine residues. We prepared variants of the cholera toxin B subunit (CtxB) with one, three, or five N-terminal glycines. In order to verify the selectivity for glycine,

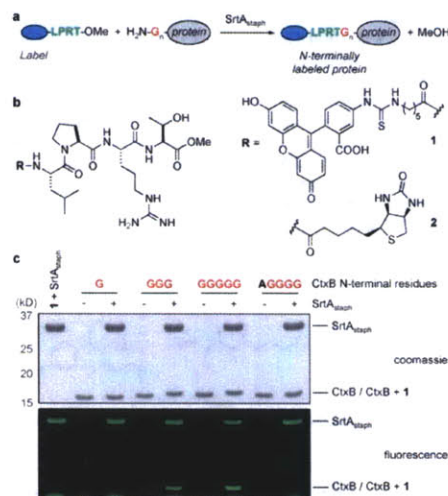


Figure 1. N-terminal labeling using SrtA<sub>staph</sub>. (a) SrtA<sub>staph</sub> catalyzes a transacylation reaction using labeled LPRT methyl esters as substrates. The labeled LPRT fragment is transferred to proteins containing N-terminal glycines in a site-specific fashion. (b) FITC (1) and biotin (2) LPRT methyl esters for N-terminal transacylation. (c) CtxB derivatives (50 μM) were treated with 500 μM 1 and 50 μM SrtA<sub>staph</sub> for 2 h at 37 °C in 50 mM Tris (pH 7.5), 150 mM NaCl, and 10 mM CaCl<sub>2</sub>. Reactions were analyzed by SDS-PAGE with visualization by coomassie staining and fluorescent gel scanning.

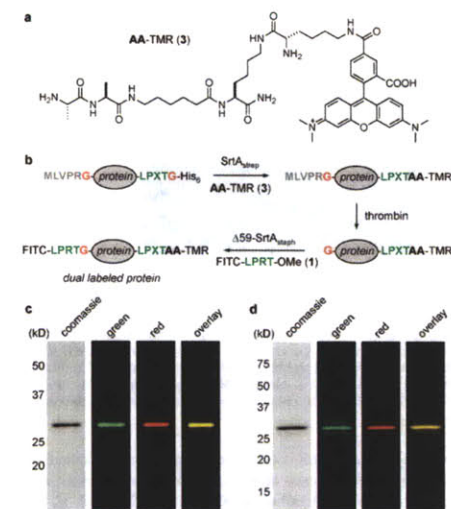
we prepared a control construct containing an N-terminal alanine residue. In the presence of SrtA<sub>staph</sub>, we observed robust labeling of G<sub>3</sub>-CtxB and G<sub>5</sub>-CtxB using 500 μM 1 for 2 h at 37 °C, with no apparent labeling of the alanine-containing control (Figure 1c). Electrospray ionization mass spectrometry (ESI-MS) revealed quantitative labeling of G<sub>3</sub>-CtxB and G<sub>5</sub>-CtxB, with no modification observed for G<sub>1</sub>-CtxB and AG<sub>4</sub>-CtxB (Figure S2). Similar experiments with biotinylated derivative 2 and G<sub>5</sub>-CtxB yielded comparable results, as verified by ESI-MS and streptavidin immunoblot (Figure S3). In all cases, residual labeling of SrtA<sub>staph</sub> itself, attributable to the formation of a covalent acyl-enzyme intermediate, was detected. N-terminal transpeptidation was also successful for two additional protein substrates, eGFP with five N-terminal glycines and UCHL3 containing a single N-terminal glycine (Figure S4).

With the ability of SrtA<sub>staph</sub> to append labels at either terminus, we pursued the possibility of installing two modifications within



the same protein. Attempts to execute this type of transformation using SrtA<sub>staph</sub> alone were unsuccessful, as intramolecular transpeptidation between N-terminal glycines and the C-terminal LPXTG motif was unavoidable in most cases. Therefore, we considered the possibility of using a second, distinct sortase, an idea that has been suggested but never reduced to practice.<sup>1,5</sup> We initially sought to use sortase B (SrtB) from either *Staph. aureus* or *Bacillus anthracis* as enzymes with recognition sequences (NPQTN and NPKTG, respectively) orthogonal to that of SrtA<sub>staph</sub>.<sup>6,7</sup> Both SrtB enzymes were easily produced in *Escherichia coli* and purified to homogeneity. We reproduced the reported in vitro enzyme activity using a FRET-based assay to measure cleavage of short peptides substrates.<sup>6,7</sup> However, to date we have failed to obtain transpeptidation with either SrtB on protein substrates modified with the appropriate recognition sequences on a time scale or with yields that compare favorably with SrtA<sub>staph</sub> (data not shown).

We ultimately arrived at a successful orthogonal strategy using SrtA<sub>strep</sub>, which recognizes the same LPXTG sequence used by SrtA<sub>staph</sub> but can accept alanine-based nucleophiles.<sup>8</sup> This leads to the formation of an LPXTA sequence at the site of ligation, a motif refractory to cleavage by SrtA<sub>staph</sub>.<sup>9</sup> This allows SrtA<sub>staph</sub> to act on the N terminus without affecting the C-terminal modification installed with SrtA<sub>strep</sub>.



**Figure 2.** Site-specific N- and C-terminal labeling using multiple sortases. (a) Tetramethylrhodamine-labeled dialanine nucleophile (3) for SrtA<sub>strep</sub>-mediated transpeptidation. (b) Strategy for the installation of discrete labels at both termini of the same protein using Δ59-SrtA<sub>staph</sub> and SrtA<sub>strep</sub>. (c, d) SDS-PAGE characterization with fluorescent gel scanning of dual-labeled (c) eGFP and (d) UCHL3.

Our final strategy for dual-terminus labeling is outlined in Figure 2b. We first synthesized a tetramethylrhodamine-labeled peptide (3) containing two N-terminal alanine residues to serve as the nucleophile for SrtA<sub>strep</sub>-mediated protein ligation (Figure 2a and Figure S5). We prepared two model substrates (eGFP and UCHL3) containing masked N-terminal glycines that are exposed only upon thrombin cleavage. Masking was required because SrtA<sub>strep</sub> was

observed to ligate both glycine and alanine nucleophiles (data not shown). Substrates also contained an LPXTG motif at the C terminus to allow a first round of labeling with SrtA<sub>strep</sub>. For both eGFP and UCHL3, C-terminal labeling using 3 and SrtA<sub>strep</sub> resulted in >90% conversion to the desired adduct, as revealed by ESI-MS (Figure S6). SrtA<sub>strep</sub> was quenched by the addition of MTSET followed by removal of His<sub>6</sub>-tagged SrtA<sub>strep</sub> using Ni-NTA. Residual 3 was then removed using a disposable desalting column. Thrombin cleavage proceeded in quantitative fashion using commercial thrombin agarose resin (Figure S6). The exposed N-terminal glycines were then labeled by treatment with 500 μM 1 and 50 μM Δ59-SrtA<sub>staph</sub><sup>10</sup> for ~1 h at 37 °C. ESI-MS of crude reaction mixtures showed the dual-labeled material as the major component, with only minor amounts of byproduct (Figure S6). A final separation by anion-exchange chromatography yielded dual-labeled eGFP and UCHL3 with excellent purity, as determined by both SDS-PAGE and ESI-MS (Figure 2c,d and Figure S6). In the case of UCHL3, we observed some additional low-intensity bands in the fluorescent gel scan (Figure 2d). However, quantitative densitometric analysis of coomassie-stained gels indicated purity in excess of 95% for both dual-labeled eGFP and UCHL3.

In summary, we have developed a strategy for placing different chemical labels at the two ends of the same polypeptide using two sortase enzymes with unique reactivities. We anticipate that this method will be applicable to the preparation of protein conjugates for refolding studies or for the construction of protein sensors, where measuring conformational changes by FRET is a common mode of detection. In more general terms, this work begins to explore the range of protein modifications that can be accessed using alternative sortases. The number of sortases that have been produced in recombinant form with retention of activity is continually increasing, and we are exploring the use of these unique enzymes as tools for protein engineering.

**Acknowledgment.** This work was supported by grants from the National Institutes of Health (R01-AI057182, R01-AI033456, R21-EB008875). J.M.A. and N.C.Y. acknowledge Clay Postdoctoral Fellowships. C.P.G. thanks the Fundacao para a Ciencia e Tecnologia (Portugal). G.-L.C. was supported by MIT's Undergraduate Research Opportunities Program (UROP). The authors thank Mark J. Banfield and colleagues for providing the expression plasmid for SrtA<sub>strep</sub> and Olaf Schneewind for providing the SrtB constructs.

**Supporting Information Available:** Full experimental details and characterization data for peptide derivatives and protein conjugates. This material is available free of charge via the Internet at <http://pubs.acs.org>.

**References**

- (1) Tsukiji, S.; Nagamune, T. *ChemBioChem* 2009, 10, 787–798.
- (2) Marraffini, L. A.; Dedent, A. C.; Schneewind, O. *Microbiol. Mol. Biol. Rev.* 2006, 70, 192–221.
- (3) Pallen, M. J.; Lam, A. C.; Antonio, M.; Dunbar, K. *Trends Microbiol.* 2001, 9, 97–102.
- (4) Yamamoto, T.; Nagamune, T. *Chem. Commun.* 2009, 1022–1024.
- (5) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nat. Chem. Biol.* 2007, 3, 707–708.
- (6) Maresco, A. W.; Chapa, T. J.; Schneewind, O. *J. Bacteriol.* 2006, 188, 8145–8152.
- (7) Muzmanian, S. K.; Ton-That, H.; Su, K.; Schneewind, O. *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99, 2293–2298.
- (8) Race, P. R.; Bentley, M. L.; Melvin, J. A.; Crow, A.; Hughes, R. K.; Smith, W. D.; Sessions, R. B.; Kehoe, M. A.; McCafferty, D. G.; Banfield, M. J. *J. Biol. Chem.* 2009, 284, 6924–6933.
- (9) Kruger, R. G.; Orvos, B.; Frankel, B. A.; Bentley, M.; Dostal, P.; McCafferty, D. G. *Biochemistry* 2004, 43, 1541–1551.
- (10) Δ59-SrtA<sub>staph</sub> is a truncated form of SrtA<sub>staph</sub> that has identical reactivity. JA902681K



Chapter 3: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificities

**Supplemental Online Material at:**

**<http://pubs.acs.org/doi/suppl/10.1021/ja902681k>**

## **Chapter 4: A straight path to circular proteins**

## Chapter 4: A straight path to circular proteins

### **A Straight Path to Circular Proteins**

(From: J. M. Antos, M. W. Popp, R. Ernst, G. L. Chew, E. Spooner, H. L. Ploegh, *J Biol Chem* **2009**, *284*, 16028)

#### **Abstract**

Folding and stability are parameters that control protein behavior. The possibility of conferring additional stability on proteins has implications for their use *in vivo*, and for their structural analysis in the laboratory. Cyclic polypeptides ranging in size from 14 to 78 amino acids occur naturally and often show enhanced resistance toward denaturation and proteolysis when compared to their linear counterparts. Native chemical ligation and intein-based methods allow production of circular derivatives of larger proteins, resulting in improved stability and refolding properties. Here we show that circular proteins can be made reversibly with excellent efficiency by means of a sortase-catalyzed cyclization reaction, requiring only minimal modification of the protein to be circularized.

#### **Introduction**

Sortases are bacterial enzymes that predominantly catalyze the attachment of surface proteins to the bacterial cell wall<sup>1-2</sup>. Other sortases polymerize pilin subunits for the construction of the covalently stabilized and covalently anchored pilus of the gram-positive bacterium<sup>3-5</sup>. The reaction catalyzed by sortase involves the recognition of short 5-residue sequence motifs, which are cleaved by the enzyme with the concomitant formation of an acyl enzyme intermediate between the active site cysteine of sortase and the carboxylate at the newly generated C-terminus of the substrate<sup>1,6-8</sup>. In many bacteria, this covalent intermediate can be resolved by nucleophilic attack from the pentaglycine sidechain in a peptidoglycan precursor, resulting in the formation of an

#### Chapter 4: A straight path to circular proteins

amide bond between the pentaglycine sidechain and the carboxylate at the cleavage site in the substrate<sup>9-10</sup>. In pilus construction, alternative nucleophiles such as lysine residues or diaminopimelic acid participate in the transpeptidation reaction<sup>3-4</sup>.

When appended near the C-terminus of proteins that are not natural sortase substrates, the recognition sequence of *Staphylococcus aureus* sortase A (LPXTG) can be used to effectuate a sortase-catalyzed transpeptidation reaction using a diverse array of artificial glycine-based nucleophiles (**Figure 4.1**). The result is efficient installation of a diverse set of moieties including lipids<sup>11</sup>, carbohydrates<sup>12</sup>, peptide nucleic acids<sup>13</sup>, biotin<sup>14</sup>, fluorophores<sup>14-15</sup>, polymers<sup>16</sup>, solid supports<sup>16-18</sup>, or peptides<sup>15,19</sup> at the C-terminus of the protein substrate. During the course of our studies to further expand sortase-based protein engineering, we were struck by the frequency and relative ease with which intramolecular transpeptidation reactions were occurring. Specifically, proteins equipped with not only the LPXTG motif, but also N-terminal glycine residues yielded covalently closed circular polypeptides (**Figure 4.1**). Similar reactivity using sortase has been described in two previous cases; however, rigorous characterization of the circular polypeptides was absent<sup>16,20</sup>. The circular proteins in these reports were observed as minor components of more complex reaction mixtures, and the cyclization reaction itself was not optimized.

Here we describe our efforts toward applying sortase-catalyzed transpeptidation to the synthesis of circular and oligomeric proteins. This method has general applicability, as illustrated by successful intramolecular reactions with three structurally unrelated

#### Chapter 4: A straight path to circular proteins

proteins. In addition to circularization of individual protein units, the multiprotein complex AAA-ATPase p97/VCP/CDC48, with six identical subunits containing the LPXTG motif and an N-terminal glycine, was found to preferentially react in daisy chain fashion to yield linear protein fusions. The reaction exploited here shows remarkable similarities to the mechanisms proposed for circularization of cyclotides, small circular proteins that have been isolated from plants<sup>21-23</sup>.

#### **Results**

*Cre recombinase.* We first noticed the presence of a circular protein product when installing a C-terminal modification onto a nonfunctional mutant of Cre recombinase containing a single N-terminal glycine residue and the requisite LPETG sequence near the C-terminus. The LPETG motif was separated from the native protein by a flexible amino acid linker (GGGGSGGGGS). Whereas installation of the label at the Cre C terminus proceeded efficiently when a triglycine nucleophile containing tetramethylrhodamine (GGG-TMR) was included, we observed a product that migrated more rapidly on SDS-PAGE when nucleophile was omitted from the reaction mixture (**Figure 4.2a**). Hydrolysis of the sortase acyl enzyme is known to proceed slowly in the absence of glycine nucleophiles<sup>19, 24-25</sup>. However, when reaction mixtures were analyzed by ESI-MS we consistently observed a protein species that differed from the mass expected for hydrolysis by ~ -18 Da (**Figure 4.2b**). This mass was consistent with intramolecular nucleophilic attack, suggesting that the single N-terminal glycine residue was serving as the nucleophile in this transformation. Ultimately, MS/MS on tryptic digests of this species showed unequivocally that it consisted of a covalently closed

#### Chapter 4: A straight path to circular proteins

circular product of Cre, with the N-terminal glycine fused exactly at the LPETG cleavage site in the expected position.

Recognizing that the LPETG motif is maintained in the cyclized Cre product, we suspected that sortase should be capable of cleaving the circular protein at this site, thus producing an equilibrium between circular and linear forms of Cre. To demonstrate this point, Cre was first incubated with sortase in the presence or absence of triglycine nucleophile (Figure 3A). A portion of the cyclized reaction mixture (**Figure 4.3a, lane 1**) was then treated with a large molar excess of triglycine nucleophile or left alone for a further 24 h (**Figure 4.3a, lanes 2-3**). Remarkably, upon treatment with exogenous nucleophile, the pre-cyclized material yielded a reaction mixture that was nearly identical to the result obtained when nucleophile was included from the very beginning of the experiment (compare lanes 3 and 4). This result provided further evidence that cyclized Cre indeed contains the expected LPETG motif at the site of covalent closure. In addition, it suggested that hydrolysis of the acyl enzyme intermediate does not effectively compete during cyclization, because the hydrolyzed material should be unable to participate in the transpeptidation reaction.

The circularization reaction observed for Cre proceeded with remarkable efficiency. Conversion was estimated to be >90% by SDS-PAGE. By taking an existing crystal structure<sup>26</sup> of the Cre protein and modeling in those residues not visible in the structure, it was clear that the N- and C-termini were located in sufficiently close proximity to permit closure without significant perturbation of the native structure (**Figure 4.3b**). We

#### Chapter 4: A straight path to circular proteins

assume that these regions possess considerable flexibility because they are not resolved in the crystal structure.

*eGFP*. Having verified the cyclization of Cre recombinase, we sought to explore the generality of this technique. To this end we generated a derivative of eGFP containing the LPETG sequence and five N-terminal glycine residues. This construct was of particular interest because inspection of the X-ray crystal structure<sup>27</sup> revealed that the N- and C-termini were positioned on the same end of the  $\beta$ -barrel, suggesting that this substrate should be ideal for cyclization (Figure 4A). Furthermore, in one of the earliest reports on the use of sortase for protein engineering a similar eGFP substrate was described and reported to cyclize in the presence of sortase<sup>16</sup>. In this instance, cyclization only proceeded in modest yield, and the putative cyclized product was produced as a mixture with higher molecular weight species assigned as oligomers of eGFP formed by intermolecular transpeptidation. Thus, to explore potential complications caused by intermolecular reactions, we studied the reaction of our eGFP construct in the presence of sortase.

In our hands, we observed clean conversion to a lower molecular weight species (>90% estimated conversion) with little to no evidence for oligomerization (Figure 4B). A higher molecular polypeptide was observed at early time points, and may represent a covalent eGFP dimer that is generated transiently over the course of the reaction. Higher molecular species, however, were only observed in trace quantities in the final reaction mixture. As in the case of Cre, evidence for circularization was provided by mass

#### Chapter 4: A straight path to circular proteins

spectral characterization of the intact circular protein and MS/MS sequencing of tryptic peptides (**Supplementary Figure 4.1**). As an additional control to demonstrate that the N-terminal glycine residue was the only nucleophile participating in intramolecular transpeptidation, we analyzed the behaviour of an eGFP derivative that lacked an N-terminal glycine. In this case, ESI-MS revealed products consistent with hydrolysis of the acyl enzyme intermediate, rather than intramolecular nucleophilic attack (**Supplementary Figure 4.1**).

Circularization has been shown to confer unique properties onto proteins when compared to the linear form<sup>28-30</sup>. In the case of GFP circularized using intein-based methods, these properties include a reduced rate of unfolding when exposed to denaturants, as well as an enhanced rate of refolding following denaturation<sup>28</sup>. We observed a similar phenomenon for eGFP circularized using sortase (**Figure 4.4c**). Circular eGFP was first separated from residual sortase A using Ni-NTA resin. This material retained fluorescence suggesting that covalent ligation of the N- and C-termini had minimal impact on the structure of this substrate. Circular and linear eGFP were then subjected to simple thermal denaturation, followed by recovery at room temperature. As shown in Figure 4C, circular eGFP regained fluorescence more rapidly than linear eGFP.

*UCHL3*. Even an internally positioned LPXTG motif was sufficient to effectuate a circularization reaction. We installed a sortase recognition site in the crossover loop of the ubiquitin C-terminal hydrolase UCHL3, and demonstrated that the continuity of the polypeptide backbone can be disrupted with concomitant installation of a covalent



#### Chapter 4: A straight path to circular proteins

modification that reports on the accuracy of cleavage and transpeptidation<sup>31</sup>. This reaction proceeds without complete loss of activity of UCHL3, indicating that even the cleaved form of UCHL3 retains its structural integrity to a significant degree<sup>31</sup>. This UCHL3 construct was prepared with an N-terminal glycine residue, and examination of the crystal structure of UCHL3<sup>32</sup> clearly showed the close apposition of the N-terminus and the crossover loop, suggesting that cyclization to yield a circular fragment containing the N-terminal portion of UCHL3 should be readily observable (**Figure 4.5a**).

As expected, in the absence of added nucleophile, the N-terminal glycine serves as a highly efficient nucleophile to yield a circular fragment that contains the N-terminal portion of UCHL3 (**Figure 4.5b**). The identity of the circular polypeptide was confirmed by MS/MS of the peptide containing the expected fusion of the N-terminal glycine residue with the new C-terminus released from the crossover loop (see **Supplemental Figure 4.2**). Cyclization was efficiently blocked if a high concentration of triglycine (GGG) was included in the reaction, generating instead the N-terminal fragment of UCHL3 transacylated onto the triglycine nucleophile (**Figure 4.5b, lane 9** and **Supplemental Figure 4.3**). Cyclization could also be reversed by adding an excess of triglycine to reaction mixtures pre-incubated with sortase to allow cyclization. This reopening reaction was observed by both SDS-PAGE and ESI-MS (**Supplemental Figure 4.3**).

To test the functional properties of cyclic UCHL3, we incubated reaction mixtures with an activity-based probe consisting of ubiquitin equipped with an electrophilic vinyl

#### Chapter 4: A straight path to circular proteins

methyl ester moiety at the C terminus (**Supplemental Figure 4.4**). Probes of this nature are able to specifically alkylate active-site cysteine residues in ubiquitin specific hydrolases such as UCHL3<sup>31,33-34</sup>. Following circularization, the active-site cysteine (C95) of UCHL3 is located in the circular N-terminal fragment, and indeed we observed covalent labeling of this fragment with a corresponding shift in apparent molecular weight consistent with the attachment of ubiquitin. This result suggests that despite cleavage of the polypeptide backbone, the circular N-terminal fragment of UCHL3 and the C-terminal portion released during transpeptidation remain associated and preserve the affinity of UCHL3 for ubiquitin. This result is consistent with previous observations from our laboratory demonstrating that covalent closure of the UCHL3 crossover loop is dispensable for enzyme activity<sup>31</sup>.

*p97*. The above examples concern single chain proteins whose termini are sufficiently close to allow covalent closure by means of the sortase-mediated transacylation reaction. Similar proximity relationships between protein termini should also be present on separate polypeptides that assemble into defined oligomeric structures. As an example, we examined *p97*, a hexameric AAA-ATPase. We generated a derivative of *p97* (G-His<sub>6</sub>-*p97*-LPSTG-XX) containing an LPSTG motif near the C terminus, and a hexahistidine tag capped by two serine residues and a single glycine at the N terminus. The structure of a *p97* trimer in the presence of ADP has been solved at 3.5 Å resolution<sup>35</sup>, with several residues from the N- and C-termini not visible (**Figure 4.6a**). When all the residues present in our modified version of *p97* were modeled onto the published trimer of *p97*, it was evident that the N- and C-termini of adjacent *p97* units were

#### Chapter 4: A straight path to circular proteins

sufficiently close enough to permit covalent crosslinking (**Figure 4.6b**). G-His<sub>6</sub>-p97-LPSTG-XX was expressed in *E. coli* and yielded the hexameric p97 ring, as assessed by gel filtration. As expected, this derivative of p97 was an excellent substrate for transpeptidation at its C terminus, allowing efficient installation of a label when incubated in the presence of sortase and GGG-TMR (**Supplemental Figure 4.5**). In contrast, a variant of p97 lacking the LPSTG sequence showed no labeling (**Supplemental Figure 4.5**). When G-His<sub>6</sub>-p97-LPSTG-XX was treated with sortase A in the absence of added nucleophile, we observed formation of a SDS-resistant ladder of polypeptides, as would be expected for intermolecular crosslinking of p97 monomers (**Figure 4.6c**). We were confident that these species arise from head-to-tail ligation of p97 because introduction of excess diglycine (GG) after oligomerization caused collapse of the higher molecular weight structures back to monomeric p97 (**Figure 4.6c, lane 5**). This suggested that the higher order aggregates are held together by newly formed LPSTG units formed from the C-terminal LPST residues of one p97 monomer and the N-terminal glycine residue of a neighboring monomer. The banding pattern observed for reopening was also nearly identical to that seen when diglycine was included from very beginning of the experiment, a scenario where installation of diglycine at the C-terminus of each p97 subunit is presumed to be the major reaction pathway (**Figure 4.6c, lane 6**). We have also been able to identify peptides consistent with intermolecular crosslinking of p97 subunits by MS/MS (**Supplemental Figure 4.6**).

## Chapter 4: A straight path to circular proteins

### **Discussion**

Cyclic proteins are an interesting class of polypeptides that often display unique properties due to covalent closure of the amide backbone<sup>36-37</sup>. While some cyclic protein derivatives occur naturally, methods for generating cyclic proteins in the laboratory provide a means for accessing cyclic versions of proteins that only occur in linear form. Intramolecular sortase-catalyzed transpeptidation provides a straightforward method for accessing these types of cyclic proteins. The transpeptidation reaction described here bears a remarkable resemblance to the proposed biosynthesis of the largest class of naturally occurring cyclic proteins, the cyclotides<sup>21-23</sup>. In both cases, linear protein precursors are cleaved by cysteine proteases to generate an acyl-enzyme intermediate that is subsequently resolved by nucleophilic attack from the N-terminus of the linear proteins to generate the cyclic product.

In this work we have explored transpeptidation reactions using four structurally diverse protein substrates. Cyclization has been confirmed for three proteins, including an example (UCHL3) utilizing an LPXTG sequence positioned in a flexible internal loop rather than near the protein C-terminus. Cyclization and oligomerization via sortase-mediated transpeptidation have been previously suggested to occur for an eGFP construct modified in a manner similar to that used here<sup>16</sup>, and for a by-product from a protein purification system where the substrate circularized appears to be sortase A itself<sup>20</sup>. In both cases, the identity of the circular products was not rigorously confirmed. Our data identify the circular or oligomeric products unambiguously by MS/MS for all substrates studied. We also find that our eGFP derivative strongly favors cyclization over

#### Chapter 4: A straight path to circular proteins

oligomerization, showing little evidence for the formation of higher order structures that might be expected by the head-to-tail ligation of termini from separate eGFP monomers. Subtle differences in the structure of the eGFP constructs cannot be overlooked as a potential cause for the observed results. For example, our eGFP is extended at the N-terminus by only five glycine residues while the construct studied by Parthasarathy *et. al.* contains an additional seventeen residues, including three N-terminal glycines<sup>16</sup>. Future work will be required to thoroughly characterize how distance relationships between protein termini favor intra- versus intermolecular transpeptidation.

With respect to protein cyclization, sortase-mediated circularization is efficient despite the potential for competing reaction pathways. In the absence of added oligoglycine nucleophile, these include hydrolysis of the acyl enzyme intermediate, reattachment of the C-terminal protein fragment that is lost upon initial cleavage of the protein substrate by sortase, or, as mentioned above, oligomerization of protein monomers in head-to-tail fashion. Even when oligoglycine nucleophile is added with the intent of blocking the cyclization pathway, millimolar concentrations are necessary to efficiently compete with cyclization. One factor that certainly must contribute to this observed preference for cyclization is the distance between protein termini. Inspection of the PDB database shows that nearly one third of proteins with known structures have their termini in rather close apposition (within 20 Å)<sup>37</sup>. The LPXTG sequence itself spans roughly 15 Å in an extended conformation, suggesting that circularization via sortase-catalyzed transpeptidation might be amenable to a significant fraction of proteins using the LPXTG sequence alone to bridge the gap between N- and C-termini. Larger distances could

#### Chapter 4: A straight path to circular proteins

simply be covered by inserting flexible amino acid spacers at either termini. We also consider it likely that the circularized version of a protein will show more restricted mobility in the segment that corresponds to the newly established LPXTG connection between its termini. This fact alone may render the circular product a comparatively worse substrate for sortase, and therefore assist in driving the transpeptidation reaction toward cyclization. As evidence for this point we have observed previously that sortase fails to cleave LPXTG motifs placed in structured loops of class I MHC molecules<sup>14</sup>.

Sortase-catalyzed transpeptidation provides an attractive alternative to existing methods for peptide and protein circularization. Chemical synthesis can provide access to circular polypeptides of modest size, with circularization of linear precursors having been achieved using native chemical ligation<sup>38-40</sup>, subtiligase<sup>41</sup>, or standard amide bond forming reactions common to solid-phase peptide synthesis<sup>40,42</sup>. For larger proteins beyond the technical capabilities of solid-phase synthesis, cyclization is most often accomplished using native chemical ligation, typically in conjunction with split-intein expression constructs<sup>28-30,43-45</sup>. When compared to the split-intein approach, the modest modification necessary to render proteins amenable to cyclization or oligomerization is certainly an attractive feature of the sortase-catalyzed process. Proteins must simply possess a sortase recognition sequence (LPXTG) either near the C terminus or in a flexible loop and an N-terminal glycine residue to act as the nucleophile. These modifications are not anticipated to have a significant impact on protein expression or function. In contrast, protein circularization by split-intein methods requires more extensive modifications of the expression construct, a necessity that may reduce protein

#### Chapter 4: A straight path to circular proteins

expression or effect protein solubility. It should be noted, however, that the number of extra amino acid residues at the site of N-to-C terminal ligation following excision of the large intein domains can be less than the five residues (LPXTG) that remain after circularization using sortase A.

The sortase-catalyzed approach also provides additional levels of control over the ensuing transpeptidation reaction. This may be particularly useful for oligomeric species, such as the p97 example described here. Specifically, our modified p97 protein (G-His<sub>6</sub>-p97-LPSTG-XX) is produced in a form that is by itself unreactive. This allows protein expression and the subsequent assembly and purification of the hexamer to be completed first, without complications caused by premature covalent oligomerization. Crosslinking is then induced by the addition of sortase after the individual subunits have been correctly positioned in the hexameric ring. The extent of transpeptidation can be further controlled by inclusion of synthetic oligoglycine nucleophiles, either during the transpeptidation reaction or after transpeptidation is complete. The latter scenario even allows cyclization to be completely reversed. Incubating circular protein products with sortase in the presence of an oligoglycine nucleophile restores linearity to the protein product, because in the course of the initial cyclization reaction, the LPXTG motif is restored. An equilibrium between closed and open forms is thus established and can be driven toward the linear state by adding a large excess of the oligoglycine nucleophile.

The implications of protein cyclization or oligomerization for protein engineering are numerous. In the case of protein oligomerization, the ability to link protein subunits held

#### Chapter 4: A straight path to circular proteins

in a defined geometry might be exploited to explore subtle changes in intersubunit interactions upon substrate engagement or recruitment of binding partners. A more detailed examination of the reaction kinetics would be required to determine, for example, whether all subunits in the hexameric ring of p97 are equally good substrates, or whether subunits that lie along the threefold axis preferentially crosslink to yield dimers. Although in the crystal structure<sup>35</sup> all of the individual subunits appear identical, it remains to be determined whether this equivalency applies in solution as well. For cyclic proteins, there is compelling evidence that demonstrates improved stability of circularized proteins when compared to their linear counterparts<sup>28-30, 37, 46</sup>. This is true for cyclic versions GFP<sup>28</sup>, b-lactamase<sup>29</sup>, and DHFR<sup>30</sup> generated using intein-based methods. The extension of protein cyclization to proteins of therapeutic value in order to improve the *in vivo* half-life has already been suggested<sup>16, 36</sup> and remains an exciting avenue for further research. Covalent closure of a protein through sortase-mediated circularization may also facilitate structural analysis of proteins whose flexible termini may interfere with crystallization.

#### **Methods**

*Synthesis of Triglycine Tetramethylrhodamine Peptide.* The structure of GGG-TMR and a detailed synthetic protocol are provided in Supplemental Text and **Supplementary Figure 4.7**.

*Cloning and Protein Expression.* Full amino acid sequences for all proteins used in this study are given in **Supplementary Figure 4.8**.



#### Chapter 4: A straight path to circular proteins

Recombinant sortase A (residues 26-206) containing an N-terminal hexahistidine tag was produced in *E. coli* as previously described<sup>8</sup>. Purified sortase A was stored in 10% (w/v) glycerol, 50 mM Tris pH 8.0, 150 NaCl at -80 °C until further use.

G-Cre-LPETG-His<sub>6</sub> was cloned into the pTriEx-1.1 Neo expression vector (Novagen) using standard molecular biology techniques. The construct contains two point mutations (M117V, E340Q) and a flexible spacer (GGGSGGGGS) inserted before the LPETG sortase recognition site. G-Cre-LPETG-His<sub>6</sub> was expressed and purified using procedures similar to those reported previously for HTNCre<sup>47</sup>. G-Cre-LPETG-His<sub>6</sub> was first transformed into Tuner (DE3) pLacI cells (Novagen) and a starter culture was grown in sterile LB media supplemented with 1% (w/v) glucose, chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL). This culture was used to inoculate a large scale culture of sterile LB containing chloramphenicol (34 µg/mL) and ampicillin (100 µg/mL). G-Cre-LPETG-His<sub>6</sub> was expressed after a 3 h induction with IPTG (0.5 mM) at 37 °C. Cells were resuspended in 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 20 mM imidazole pH 8.0. The suspension was adjusted to 50 µg/mL DNaseI, 460 µg/mL lysozyme, and 1 mM MgCl<sub>2</sub> and incubated at 4 °C for 1.5 h. The suspension was then sonicated and centrifuged. The clarified lysate was then treated with Ni-NTA agarose (Qiagen) for 1 h at 4 °C. The resin was washed with 12 column volumes of 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 20 mM imidazole pH 8.0 followed by 4 column volumes of 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 30 mM imidazole pH 8.0. The protein was then eluted with 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 300 mM imidazole pH 8.0. The purified protein was then dialyzed first against 20

#### Chapter 4: A straight path to circular proteins

mM Tris pH 7.5, 500 mM NaCl followed by 50% (w/v) glycerol, 20 mM Tris pH 7.5, 500 mM NaCl. G-Cre-LPETG-His<sub>6</sub> was then passed through a 0.22 μM filter to remove minor precipitation and stored at 4 °C.

G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> was prepared from a previously reported eGFP construct lacking the five N-terminal glycine residues using a Quickchange<sup>®</sup> II Site-Directed Mutagenesis Kit (Stratagene) and produced in *E. coli* using reported procedures<sup>11</sup>. Purified G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> was buffer exchanged into 20 mM Tris pH 8.0, 150 mM NaCl and stored at 4 °C.

UHL3 with the sortase recognition sequence (LPETG) substituted for amino acids 159-163 was cloned and produced in *E. coli* as described previously<sup>31</sup>.

Human p97 (806 aa) was PCR amplified and cloned via the *NdeI* and *HindIII* restriction sites into a pET28a+ expression vector (Novagen) to yield the G-His<sub>6</sub>-97 construct. G-His<sub>6</sub>-p97-LPSTG-XX was generated by introducing two point mutations (G782L and Q785T) and a stop codon at position 791 using Quickchange<sup>®</sup> mutagenesis (Stratagene). Recombinant p97 was expressed at 30 °C in *E. coli* after induction for 3 h with 0.5 mM IPTG. Cells were resuspended in buffer A (50 mM Tris pH 8.0, 300 mM NaCl, 5% (w/v) glycerol, 20 mM imidazole and 7.1 mM β-ME), adjusted to 15 μg/mL lysozyme and 10 μg/mL DNase I, and lysed by two passes through a French pressure cell at 1200 psi. After centrifugation for 30 min at 40,000 x g the supernatant was bound to Ni-Sepharose resin (GE Healthcare). After washing the resin with 20 column volumes of buffer A, p97

#### Chapter 4: A straight path to circular proteins

was eluted with buffer A containing 250 mM imidazole. Hexameric rings of p97 were further purified on a Superdex 200 HR 16/60 column (GE Healthcare) using 25 mM Tris pH 8.0, 150 mM KCl, 2.5 mM MgCl<sub>2</sub>, 5% (w/v) glycerol as the mobile phase. The purified protein was snap frozen and stored at -80 °C.

*Circularization and Intermolecular Transpeptidation.* Transpeptidation reactions were performed by combining the necessary proteins/reagents at the specified concentrations in the presence of sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) and incubating at 37 °C for the times indicated. Diglycine (GG) and triglycine (GGG) peptides were purchased from Sigma. Reactions were halted by the addition of reducing Laemmli sample buffer and analyzed by SDS-PAGE. Gels were visualized by staining with coomassie blue. Fluorescence was visualized on a Typhoon 9200 Imager (GE Healthcare). Crude reactions were also diluted into either 0.1% formic acid or water for ESI-MS analysis. ESI-MS was performed on a Micromass LCT mass spectrometer (Micromass<sup>®</sup> MS Technologies, USA) and a Paradigm MG4 HPLC system equipped with a HTC PAL autosampler (Michrom BioResources, USA) and a Waters Symmetry 5 μm C8 column (2.1 x 50 mm, MeCN:H<sub>2</sub>O (0.1% formic acid) gradient mobile phase, 150 μL/min).

*Purification and Refolding of eGFP.* G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> (50 μM) was circularized by treatment with sortase A (50 μM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 24 h at 37 °C. The reaction was run on 750 μL scale. The entire reaction was then diluted into 10 mL of 20 mM Tris, 500 mM NaCl, and 20 mM

#### Chapter 4: A straight path to circular proteins

imidazole pH 8.0. This solution was then applied to a column consisting of 2 mL of Ni-NTA agarose (Qiagen) pre-equilibrated with 20 mM Tris, 500 mM NaCl, and 20 mM imidazole pH 8.0. The flow through was then concentrated and buffer exchanged in 20 mM Tris, 150 mM NaCl pH 8.0 using a NAP<sup>TM</sup> 5 Sephadex<sup>TM</sup> column (GE Healthcare). The concentrations of circular eGFP and linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> were estimated by UV-Vis spectroscopy using the absorbance of eGFP at 488 nm (extinction coefficient 55,900 M<sup>-1</sup>cm<sup>-1</sup>)<sup>48</sup>. Circular and linear eGFP (40 μL of 18 μM solutions) were placed in 1.5 mL microcentrifuge tubes and denatured by heating to 90 °C for 5 min. Samples were then incubated at room temperature in the dark for the times indicated. Fluorescent images were acquired using a UV gel documentation system (UVP Laboratory Products).

*Reaction of Cyclic UCHL3 with Activity-Based Ubiquitin Probe.* UCHL3 (30 μM) was incubated with sortase A (150 μM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) in the presence or absence of 90 mM GGG peptide (Sigma) on a 25 μL scale at 37 °C for 3 hours. Ten microliters was withdrawn and diluted with 10 μL of labeling buffer (100 mM Tris pH 7.5, 150 mM NaCl). Hemagglutinin-epitope tagged ubiquitin vinyl methyl ester (HA-UbVME) (4 μg) was added as well as 1 mM DTT and incubated at room temperature for 1 hour. Reactions were then separated on an SDS-PAGE gel and visualized by coomassie staining or α-HA immunoblot (supplemental Fig. S4). HA-UbVME was prepared following published protocols<sup>33</sup>.

*MS/MS Sequencing of Proteolytic Fragments from Circular Proteins.* Prior to MS/MS analysis, circular eGFP and Cre were separated from sortase A by RP-HPLC using an

#### Chapter 4: A straight path to circular proteins

Agilent 1100 Series HPLC system equipped with a Waters Delta Pak 5  $\mu\text{m}$ , 100  $\text{\AA}$  C18 column (3.9 x 150 mm, MeCN:H<sub>2</sub>O gradient mobile phase containing 0.1% trifluoroacetic acid, 1 mL/min). Fractions containing the circular proteins were pooled and subjected to trypsin digestion. Crude transpeptidation reactions containing circular UCHL3 were separated by SDS-PAGE followed by coomassie staining. The band corresponding to circular UCHL3 was excised and digested with Glu-C. Crude transpeptidation reactions containing dimeric p97 were separated by SDS-PAGE followed by coomassie staining. The transpeptidation reaction used for this purpose was incubated for only 2 h and therefore contains less oligomerization than that seen after an overnight incubation (see **Supplementary Figure 4.6**). The band corresponding to dimeric p97 was excised and digested with chymotrypsin. For all protein substrates, the peptides generated from proteolytic digestion were extracted and concentrated for analysis by RP-HPLC and tandem mass spectrometry. RP-HPLC was carried out on a Waters NanoAcquity HPLC system with a flow rate of 250 nL/min and mobile phases of 0.1% formic acid in water and 0.1% formic acid in acetonitrile. The gradient used was isocratic 1% acetonitrile for 1 min followed by 2% acetonitrile per minute to 40% acetonitrile. The analytical column was 0.075  $\mu\text{m}$  x 10 cm with the tip pulled to 0.005  $\mu\text{m}$  and self-packed with 3  $\mu\text{m}$  Jupiter C18 (Phenomenex). The column was interfaced to a Thermo LTQ linear ion trap mass spectrometer in a nanospray configuration and data was collected in full scan mode followed by MS/MS analysis in a data dependant manner. The mass spectral data was database searched using SEQUEST.

#### Chapter 4: A straight path to circular proteins

*Construction of Molecular Models.* Molecular models were generated from published crystal structures (PDB IDs: 1kbu, 1gfl, 1xd3, and 3cfl)<sup>26-27, 32, 35</sup>. N- and C-terminal residues were added using Coot 0.5<sup>49</sup>. Protein termini were repositioned using the Auto Sculpting function in MacPyMOL (DeLano Scientific LLC). Residues visible in the published crystal structures were not moved during positioning of the extended N and C termini. All protein images in this manuscript were generated using MacPyMOL.

#### References

1. Marraffini, L.A., Dedent, A.C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* **70**, 192-221 (2006).
2. Paterson, G.K. & Mitchell, T.J. The biology of Gram-positive sortase enzymes. *Trends Microbiol* **12**, 89-95 (2004).
3. Budzik, J.M. et al. Amide bonds assemble pili on the surface of bacilli. *Proc Natl Acad Sci U S A* **105**, 10215-10220 (2008).
4. Budzik, J.M., Oh, S.Y. & Schneewind, O. Cell Wall Anchor Structure of BcpA Pili in *Bacillus anthracis*. *J Biol Chem* **283**, 36676-36686 (2008).
5. Falker, S. et al. Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol Microbiol* **70**, 595-607 (2008).
6. Kruger, R.G. et al. Analysis of the substrate specificity of the *Staphylococcus aureus* sortase transpeptidase SrtA. *Biochemistry* **43**, 1541-1551 (2004).
7. Pallen, M.J., Lam, A.C., Antonio, M. & Dunbar, K. An embarrassment of sortases - a richness of substrates? *Trends Microbiol* **9**, 97-102 (2001).
8. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci U S A* **96**, 12424-12429 (1999).
9. Schneewind, O., Fowler, A. & Faull, K.F. Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* **268**, 103-106 (1995).
10. Ton-That, H. & Schneewind, O. Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction. *J Biol Chem* **274**, 24316-24320 (1999).
11. Antos, J.M., Miller, G.M., Grotenbreg, G.M. & Ploegh, H.L. Lipid Modification of Proteins through Sortase-Catalyzed Transpeptidation. *J. Am. Chem. Soc.* **130**, 16338-16343 (2008).
12. Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V.K. & Roy, R.P. Peptide-Sugar Ligation Catalyzed by Transpeptidase Sortase: A Facile Approach to Neoglycoconjugate Synthesis. *J. Am. Chem. Soc.* **130**, 2132-2133 (2008).

Chapter 4: A straight path to circular proteins

13. Pritz, S. et al. Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. *J. Org. Chem.* **72**, 3909-3912 (2007).
14. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**, 707-708 (2007).
15. Tanaka, T., Yamamoto, T., Tsukiji, S. & Nagamune, T. Site-specific protein modification on living cells catalyzed by Sortase. *Chembiochem* **9**, 802-807 (2008).
16. Parthasarathy, R., Subramanian, S. & Boder, E.T. Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation. *Bioconjugate Chem.* **18**, 469-476 (2007).
17. Chan, L. et al. Covalent Attachment of Proteins to Solid Supports and Surfaces via Sortase-Mediated Ligation. *PLoS ONE* **2**, e1164 (2007).
18. Clow, F., Fraser, J.D. & Proft, T. Immobilization of proteins to biacore sensor chips using Staphylococcus aureus sortase A. *Biotechnol Lett* (2008).
19. Mao, H., Hart, S.A., Schink, A. & Pollok, B.A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **126**, 2670-2671 (2004).
20. Mao, H. A self-cleavable sortase fusion for one-step purification of free recombinant proteins. *Protein Expr. Purif.* **37**, 253-263 (2004).
21. Gillon, A.D. et al. Biosynthesis of circular proteins in plants. *Plant J* **53**, 505-515 (2008).
22. Saska, I. & Craik, D.J. Protease-catalysed protein splicing: a new post-translational modification? *Trends Biochem Sci* **33**, 363-368 (2008).
23. Saska, I. et al. An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *J Biol Chem* **282**, 29721-29728 (2007).
24. Huang, X. et al. Kinetic mechanism of Staphylococcus aureus sortase SrtA. *Biochemistry* **42**, 11307-11315 (2003).
25. Ton-That, H., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J. Biol. Chem.* **275**, 9876-9881 (2000).
26. Martin, S.S., Pulido, E., Chu, V.C., Lechner, T.S. & Baldwin, E.P. The Order of Strand Exchanges in Cre-LoxP Recombination and its Basis Suggested by the Crystal Structure of a Cre-LoxP Holliday Junction Complex. *J. Mol. Biol.* **319**, 107-127 (2002).
27. Yang, F., Moss, L.G. & Phillips, G.N. The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14**, 1246-1251 (1996).
28. Iwai, H., Lingel, A. & Pluckthun, A. Cyclic green fluorescent protein produced in vivo using an artificially split PI-Pful intein from Pyrococcus furiosus. *J Biol Chem* **276**, 16548-16554 (2001).
29. Iwai, H. & Pluckthun, A. Circular beta-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett* **459**, 166-172 (1999).
30. Scott, C.P., Abel-Santos, E., Wall, M., Wahnou, D.C. & Benkovic, S.J. Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci US A* **96**, 13638-13643 (1999).

Chapter 4: A straight path to circular proteins

31. Popp, M.W., Artavanis-Tsakonas, K. & Ploegh, H.L. Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations. *J. Biol. Chem.* **284**, 3593-3602 (2008).
32. Misaghi, S. et al. Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. *J. Biol. Chem.* **280**, 1512-1520 (2005).
33. Borodovsky, A. et al. Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme. *Chemistry & Biology* **9**, 1149-1159 (2002).
34. Love, K.R., Catic, A., Schlieker, C. & Ploegh, H.L. Mechanisms, biology and inhibitors of deubiquitinating enzymes. *Nature Chemical Biology* **3**, 697-705 (2007).
35. Davies, J.M., Brunger, A.T. & Weis, W.I. Improved structures of full-length p97, an AAA ATPase: Implications for mechanisms of nucleotide-dependent conformational change. *Structure* **16**, 715-726 (2008).
36. Craik, D.J. Seamless proteins tie up their loose ends. *Science* **311**, 1563-1564 (2006).
37. Trabi, M. & Craik, D.J. Circular proteins--no end in sight. *Trends Biochem Sci* **27**, 132-138 (2002).
38. Camarero, J.A. & Muir, T.W. Chemoselective backbone cyclization of unprotected peptides. *Chem. Commun.*, 1369-1370 (1997).
39. Camarero, J.A., Pavel, J. & Muir, T.W. Chemical synthesis of a circular protein domain: Evidence for folding-assisted cyclization. *Angewandte Chemie-International Edition* **37**, 347-349 (1998).
40. Daly, N.L., Love, S., Alewood, P.F. & Craik, D.J. Chemical synthesis and folding pathways of large cyclic polypeptides: studies of the cystine knot polypeptide kalata B1. *Biochemistry* **38**, 10606-10614 (1999).
41. Jackson, D.Y., Burnier, J.P. & Wells, J.A. Enzymatic Cyclization of Linear Peptide Esters Using Subtiligase. *J. Am. Chem. Soc.* **117**, 819-820 (1995).
42. Hartgerink, J.D., Granja, J.R., Milligan, R.A. & Ghadiri, M.R. Self-assembling peptide nanotubes. *J. Am. Chem. Soc.* **118**, 43-50 (1996).
43. Camarero, J.A. et al. Rescuing a destabilized protein fold through backbone cyclization. *J Mol Biol* **308**, 1045-1062 (2001).
44. Camarero, J.A., Kimura, R.H., Woo, Y.H., Shekhtman, A. & Cantor, J. Biosynthesis of a fully functional cyclotide inside living bacterial cells. *Chembiochem* **8**, 1363-1366 (2007).
45. Evans, T.C., Jr. et al. Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. *J Biol Chem* **275**, 9091-9094 (2000).
46. Colgrave, M.L. & Craik, D.J. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* **43**, 5965-5975 (2004).
47. Peitz, M. et al. Enhanced purification of cell-permeant Cre and germline transmission after transduction into mouse embryonic stem cells. *Genesis* **45**, 508-517 (2007).
48. Tsien, R.Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **67**, 509-544 (1998).



#### Chapter 4: A straight path to circular proteins

49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **60**, 2126-2132 (2004).

#### **Footnotes**

The abbreviations used are: GGG-TMR, triglycine tetramethylrhodamine peptide; eGFP, enhanced green fluorescent protein; UCHL3, ubiquitin C-terminal hydrolase L3;  $\beta$ -ME,  $\beta$ -mercaptoethanol; GG, diglycine peptide; GGG, triglycine peptide; ESI-MS, electrospray ionization mass spectrometry; RP-HPLC, reversed-phase high performance liquid chromatography; HA-UbVME, hemagglutinin epitope-tagged ubiquitin vinyl methyl ester; DTT, dithiothreitol; MS/MS, tandem mass spectrometry; ADP, adenosine diphosphate; MHC, major histocompatibility complex; DHFR, dihydrofolate reductase.

#### **Figure Legends**

##### **Figure 4.1. Scheme**

Protein substrates equipped with a sortase A recognition sequence (LPXTG) can participate in intermolecular transpeptidation with synthetic oligoglycine nucleophiles (left) or intramolecular transpeptidation if an N-terminal glycine residue is present (right).

##### **Figure 4.2. Cyclization of Cre recombinase.**

(a) G-Cre-LPETG-His<sub>6</sub> (50  $\mu$ M) was incubated with sortase A (50  $\mu$ M) in the presence or absence of fluorescent GGG-TMR (10 mM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 21.5 h at 37 °C. SDS-PAGE revealed the expected C-terminal transpeptidation product when GGG-TMR was included, whereas

omission of the triglycine nucleophile resulted in clean conversion to a unique protein species with a lower apparent molecular weight. Cartoon representations to the right of the gel indicate the topology of the protein species produced by transeptidation. (b) ESI-MS of linear G-Cre-LPETG-His<sub>6</sub> and circular Cre formed by intramolecular transeptidation. (c) MS/MS spectrum of a tryptic fragment of circular Cre showing the ligation of the N-terminal residues (GEFAPK) to the C-terminal LPET motif. Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. Only the most prominent daughter ions have been labeled in the MS/MS spectrum.

**Figure 4.3. Cyclization of Cre is reversible.**

(a) G-Cre-LPETG-His<sub>6</sub> (50 μM) was circularized by treatment with sortase A (50 μM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 21.5 h at 37 °C (lane 1). This reaction mixture was then treated with 10 mM GGG-TMR (lane 3) or simply incubated for an additional 24 h at 37 °C (lane 2). All reactions were analyzed by SDS-PAGE with visualization by coomassie staining or in-gel fluorescence. For comparison, a C-terminal labeling reaction performed using 10 mM GGG-TMR without prior cyclization of the Cre substrate (lane 4) and a sample of linear G-Cre-LPETG-His<sub>6</sub> incubated without sortase A or nucleophile (lane 5) are included.

(b) Molecular model of Cre recombinase monomer (generated from PDB ID: 1kbu)<sup>26</sup> showing the proximity relationship between the N- and C-termini. The N-terminal glycine residue is highlighted in red and the C-terminal LPET residues are shown in green.

**Figure 4.4. Circularization of eGFP.**

(a) Molecular model of eGFP (generated from PDB ID: 1gfl)<sup>27</sup> showing the proximity relationship between the N- and C-termini. The N-terminal glycine residue is highlighted in red and the C-terminal LPET residues are shown in green.

(b) G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> (50 μM) was circularized by treatment with sortase A (50 μM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 24 h at 37 °C. Cartoon representations to the right of the gel indicate the topology of the protein species produced by transpeptidation.

(c) Circular eGFP (C) recovers fluorescence more rapidly than linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> (L) following thermal denaturation for 5 min at 90 °C. SDS-PAGE analysis confirmed the purity and concentration of the circular eGFP (C) and linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> (L) samples used for refolding.

**Figure 4.5. UCHL3 with an internally positioned LPETG motif is circularized by sortase A.**

(a) Molecular model of human UCHL3 (generated from PDB ID: 1xd3)<sup>32</sup> showing the active-site crossover loop bearing an LPETG substitution (LPET residues shown in green, G residue shown in red). The N-terminal glycine residue that serves as the nucleophile for intramolecular transpeptidation is highlighted in red.

(b) UCHL3 (30 μM) bearing an LPETG substitution in the active-site crossover loop was incubated with sortase A (150 μM) in the absence or presence of GGG nucleophile (90 mM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for

the indicated times at 37 °C and analyzed by SDS-PAGE, followed by coomassie staining. Cartoon representations to the right of the gel indicate the topology of the protein species produced by transpeptidation.

**Figure 4.6. The positions of the N- and C-termini in the p97 hexamer are suitable for intermolecular crosslinking through sortase-catalyzed transpeptidation.**

(a) Molecular model of p97 trimer (generated from PDB ID: 3cf1)<sup>35</sup> showing the relative position of p97 monomers in the hexameric ring. The visible N- and C-termini from the published p97 trimer structure are indicated in red and green, respectively.

(b) Molecular model of G-His<sub>6</sub>-p97-LPSTG-XX showing the proximity relationship between N- and C-termini in adjacent p97 monomers. N and C terminal residues not visible in the published crystal structure have been modeled onto the existing structure. N terminal glycine residues are shown in red and the C terminal LPST residues are shown in green. For clarity, the C-terminal domains of the outer monomers are hidden, as is the N-terminal domain of the central monomer.

(c) G-His<sub>6</sub>-p97-LPSTG-XX (1.5 mg/mL) was incubated with sortase A (30 μM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for the times indicated at 37 °C. After 22 h, diglycine (GG) was added (100 mM final concentration), resulting in disappearance of the covalent oligomers (lane 5). For comparison, a control reaction containing 100 mM GG from the outset of the experiment is shown in lane 6.

Figure 4.1

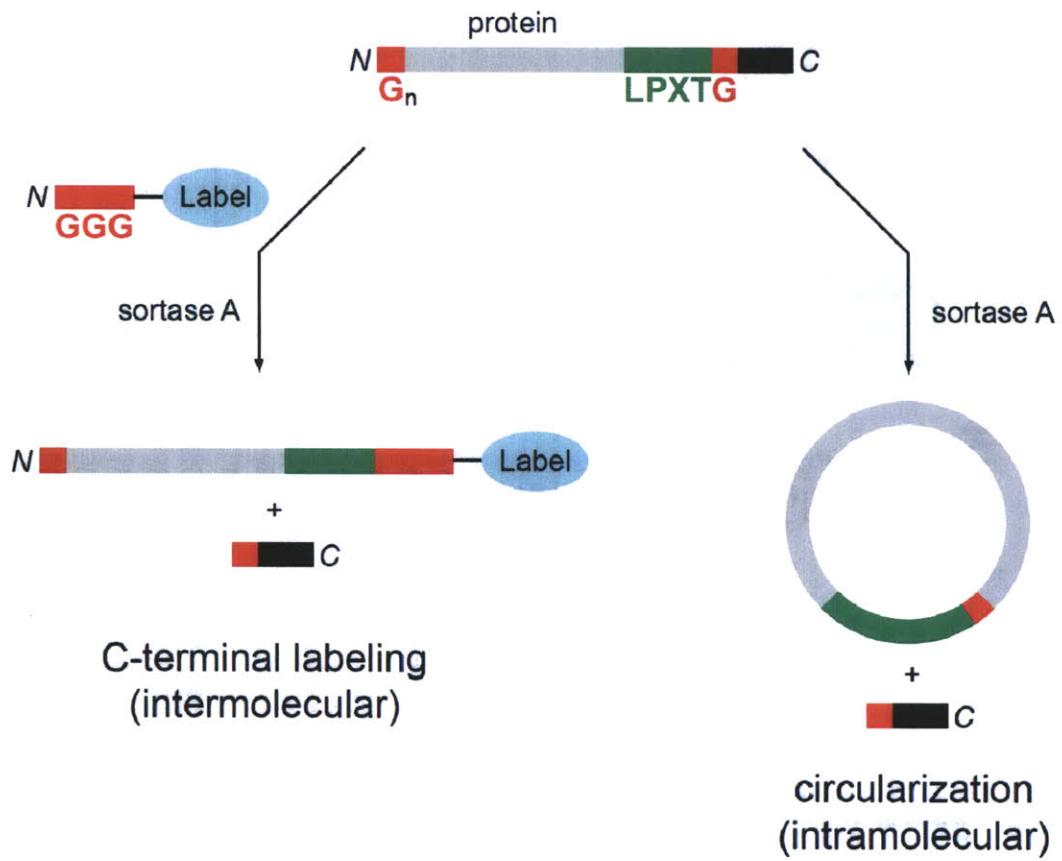


Figure 4.2

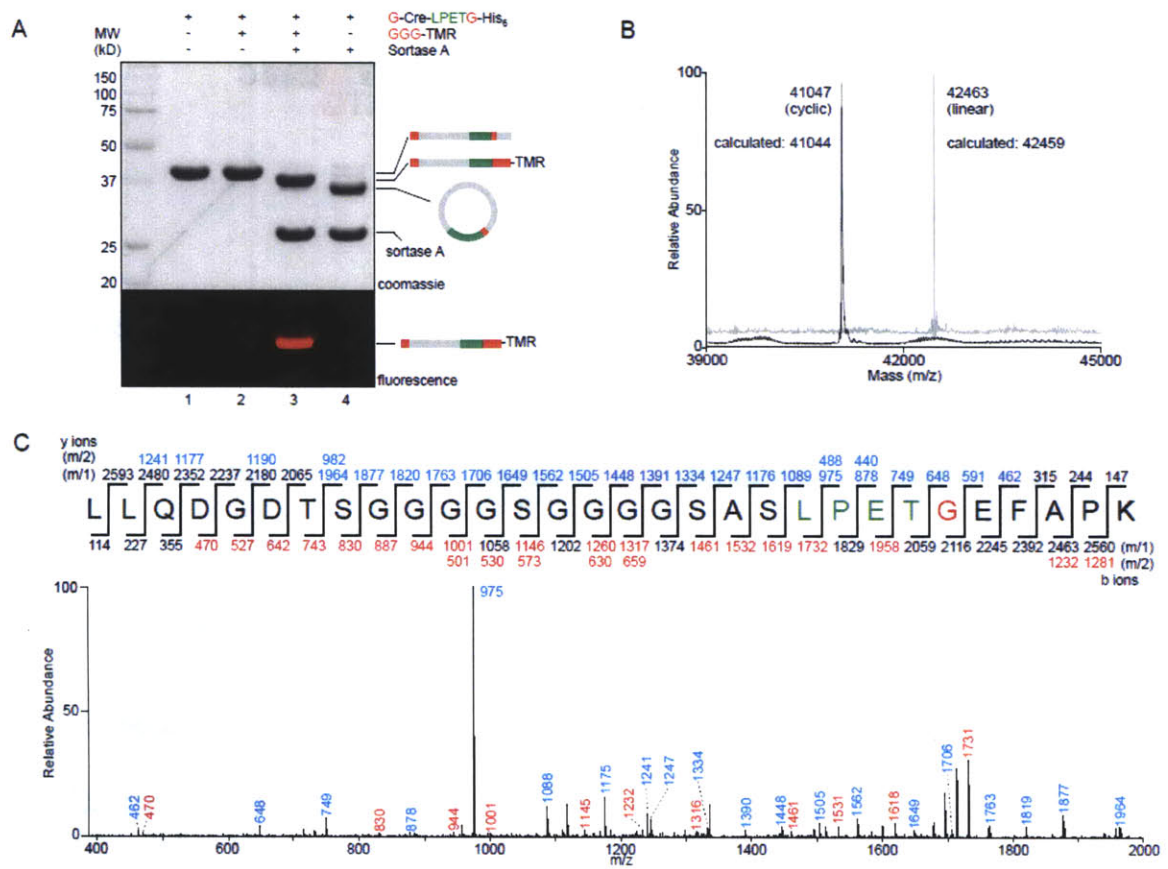


Figure 4.3

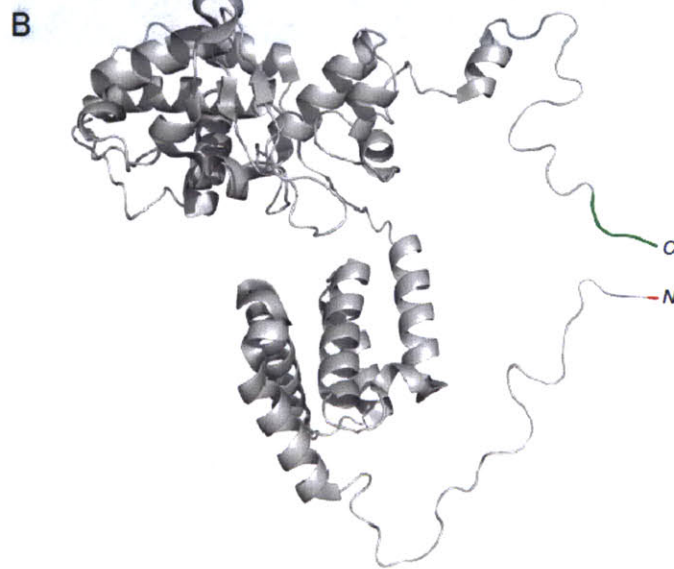
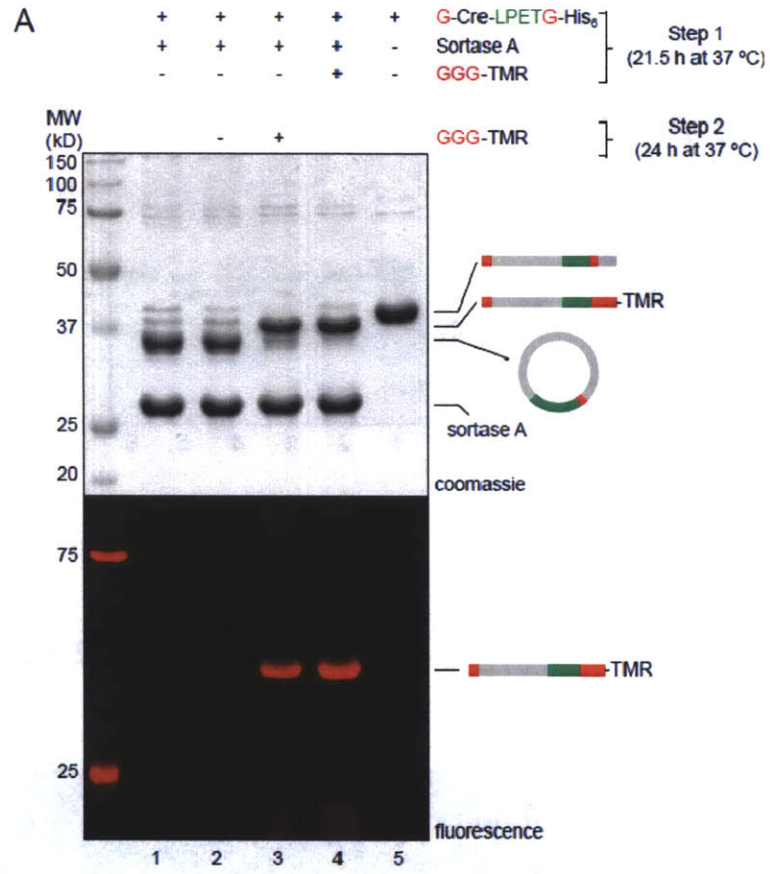


Figure 4.4

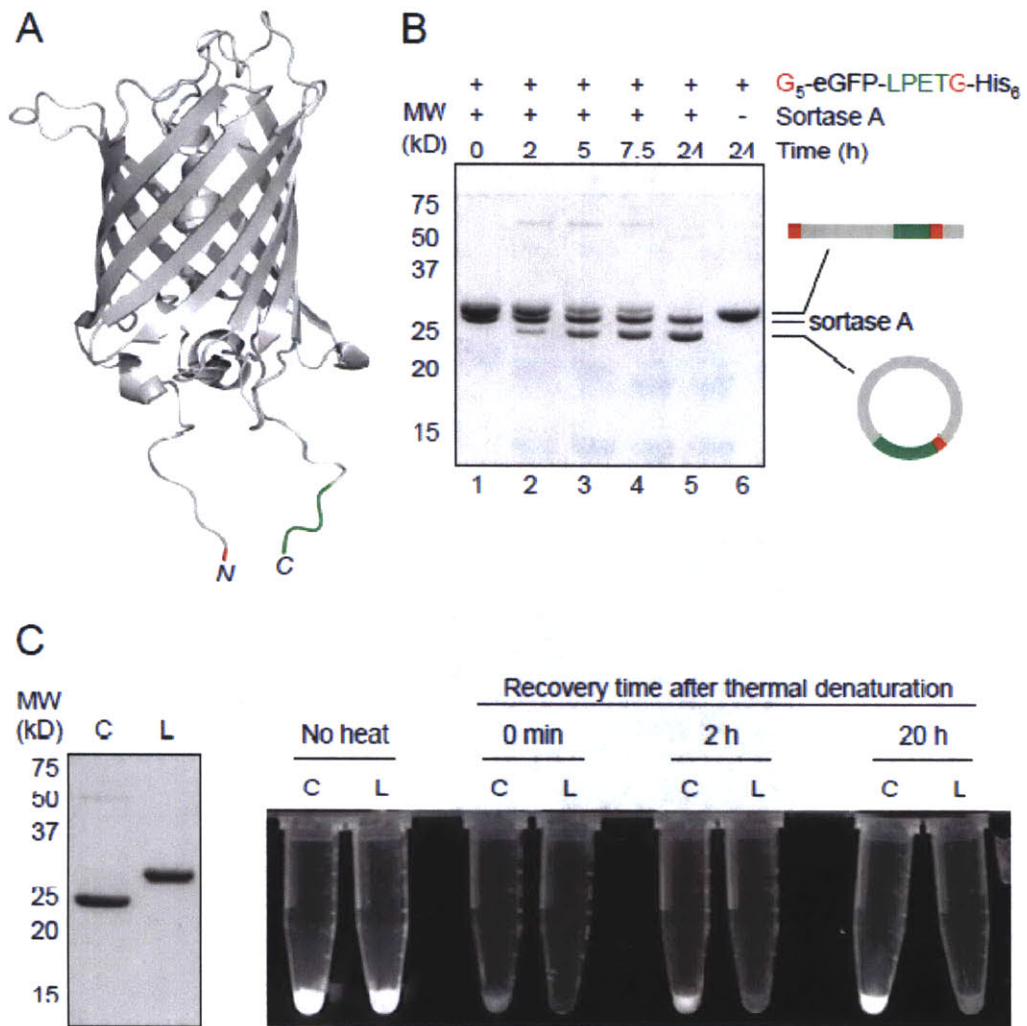




Figure 4.5

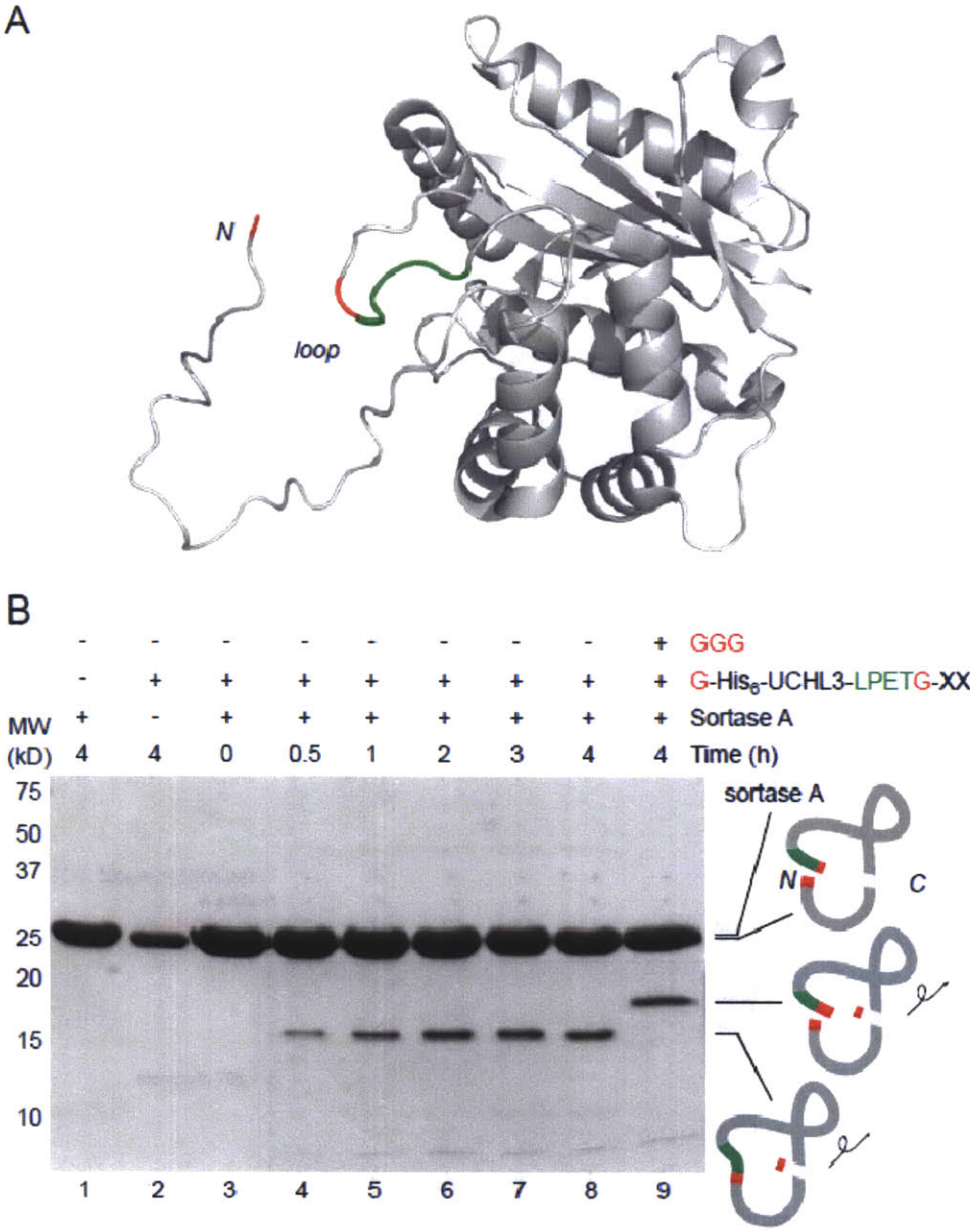
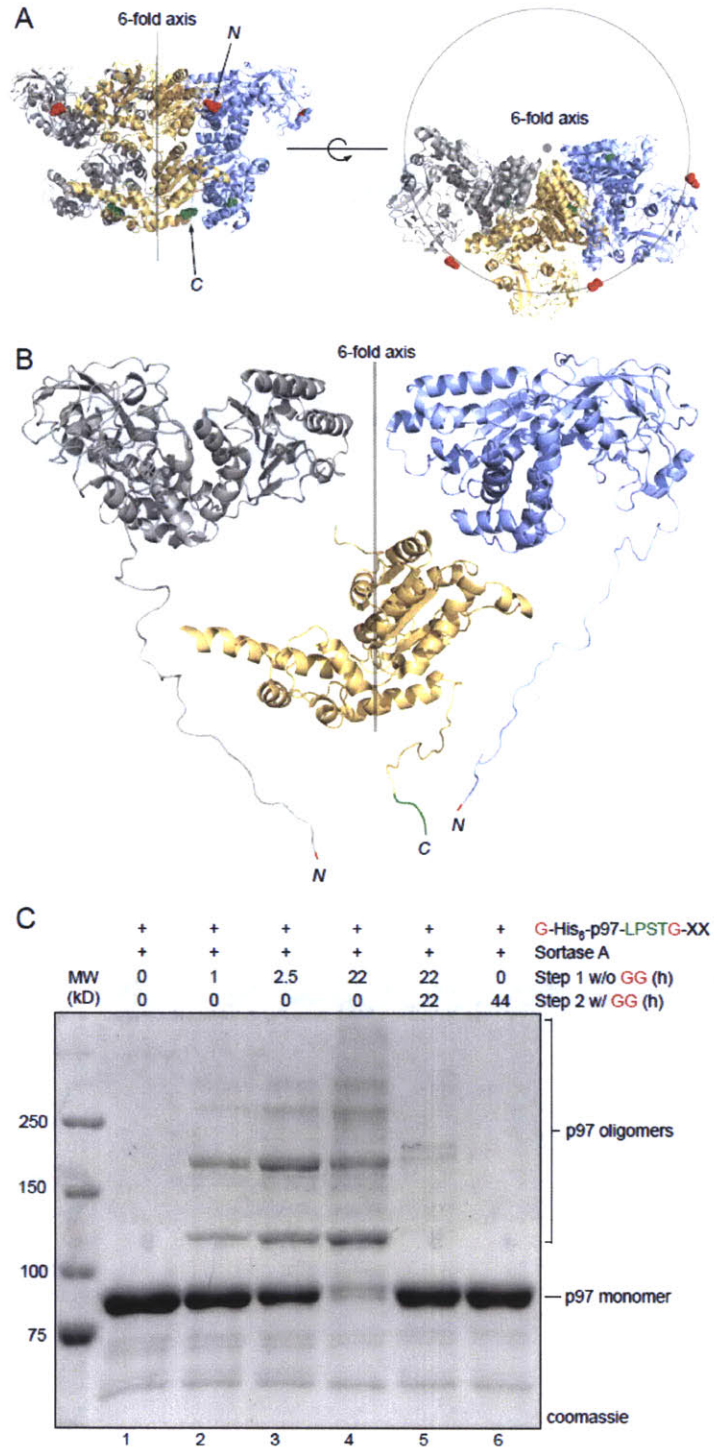


Figure 4.6



**Supplementary Figure Legends**

**Supplementary Figure 4.1. Characterization of cyclic eGFP.**

(a) ESI-MS of linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> and circular eGFP formed by intramolecular transpeptidation.

(b) A variant of eGFP lacking N-terminal glycine residues yields a mixture of unreacted starting material and hydrolysis when treated with sortase A. Conditions: eGFP-LPETG-His<sub>6</sub> (36 μM) and sortase A (50 μM) in (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 23 hours at 37 °C. This preparation of eGFP-LPETG-His<sub>6</sub> consisted of a mixture in which a fraction of the protein had the N-terminal Met residue cleaved during the course of protein expression to reveal an N-terminal Val residue.

(c) MS/MS spectrum of a tryptic fragment of circular eGFP showing ligation of the N-terminal glycine residues (shown in red) to the C-terminal LPET residues (shown in green). Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. The full amino acid sequence of linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> is shown at the top.

**Supplementary Figure 4.2. Characterization of cyclic UCHL3 by MS/MS analysis unambiguously confirms the structure of the intramolecular transpeptidation product.**

The circular UCHL3 fragment was resolved by SDS-PAGE and the band corresponding to the putative circularization product was excised and digested with Glu-C. The resulting peptides were analyzed by LC-MS/MS. A peptide representing the junction

*Chapter 4: A straight path to circular proteins*

between the UCHL3 N-terminus and the LPETG motif was identified by LC-MS/MS analysis. Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. The full amino acid sequence of linear G-His<sub>6</sub>-UCHL3-LPETG-XX is shown at the top.

**Supplementary Figure 4.3. Intermolecular versus intramolecular transpeptidation for UCHL3.**

(a) UCHL3 cyclization is reversible. UCHL3 (30  $\mu$ M) was circularized by the addition of sortase A (150  $\mu$ M) for 3 hours. Subsequently, GGG (90 mM) was added to each reaction and incubated for the indicated times. Samples were analyzed by SDS-PAGE (top) or ESI-MS (bottom).

(b) A large molar excess of GGG nucleophile is required to compete with the intramolecular UCHL3 circularization. UCHL3 (30  $\mu$ M) and sortase A (150  $\mu$ M) were incubated with the indicated concentrations of GGG nucleophile in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 3 hours at 37 °C, and analyzed by SDS-PAGE followed by coomassie staining.

**Supplementary Figure 4.4. Circular UCHL3 is competent to react with a ubiquitin suicide probe.**

UCHL3 (30  $\mu$ M) was first circularized (lane 5) by treatment with sortase A (150  $\mu$ M) or incubated with 90 mM GGG and sortase A (150  $\mu$ M) to yield the linear transpeptidation product (lane 7) for 3 hours at 37 °C. Samples were then withdrawn, diluted two-fold in

#### Chapter 4: A straight path to circular proteins

labeling buffer (100 mM Tris, pH 8.0, 150 mM NaCl) and incubated with 4  $\mu$ g of HA-UbVME for an additional hour at 25 °C. Reactions were quenched with sample buffer, analyzed by 12.5% SDS-PAGE, and either stained with coomassie (top) or transferred to nitrocellulose for anti-HA immunoblot (bottom).

#### **Supplementary Figure 4.5. C-terminal labeling of p97.**

(a) G-His<sub>6</sub>-p97-LPSTG-XX (1.5 mg/mL) and G-His<sub>6</sub>-p97 (1.5 mg/mL) were incubated with sortase A (50  $\mu$ M) and GGG-TMR (1 mM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 2.5 h at 37 °C. SDS-PAGE revealed the expected C-terminal transpeptidation product for G-His<sub>6</sub>-p97-LPSTG-XX only.

#### **Supplementary Figure 4.6. Characterization of Dimeric p97**

G-His<sub>6</sub>-p97-LPSTG-XX was incubated with sortase A in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 2 hours at 37 °C. The extent of oligomerization at this early time point was much less than that seen after overnight incubation. The reaction mixture was then separated by SDS-PAGE. The band corresponding to dimeric p97 was excised, digested with chymotrypsin, and the resulting peptides analyzed by MS/MS. A peptide fragment showing the ligation of the N-terminal glycine residue (shown in red) to the C-terminal LPST residues (shown in green) was identified. Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. Only the most prominent daughter ions have been labeled in the MS/MS

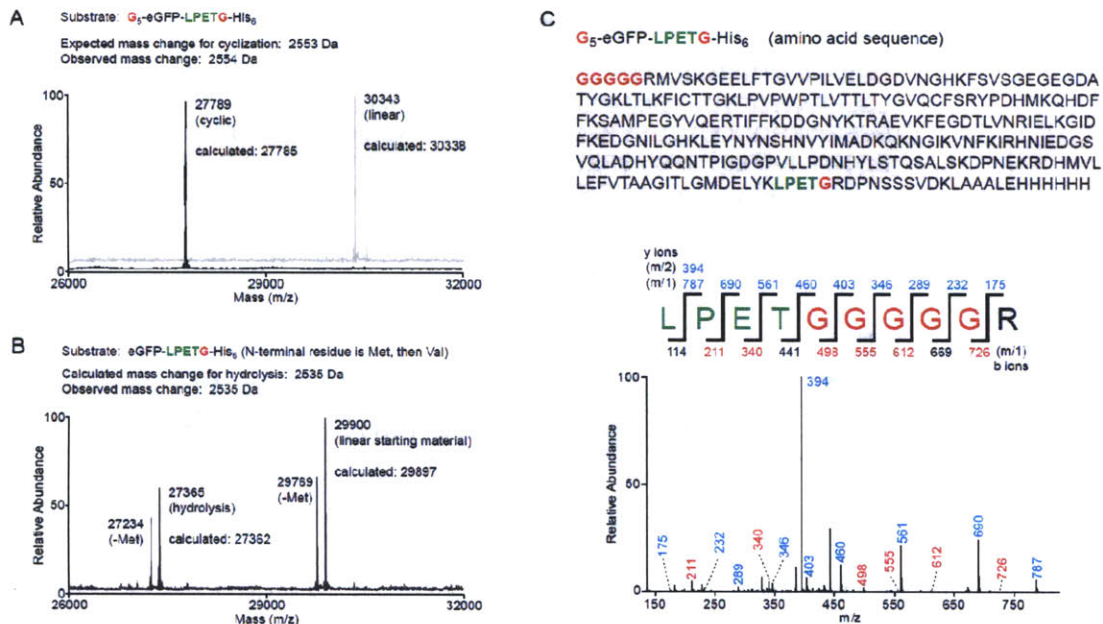
*Chapter 4: A straight path to circular proteins*

spectrum. The full amino acid sequence of linear G-His<sub>6</sub>-p97-LPSTG-XX is shown at the top.

**Supplementary Figure 4.7. Synthesis of GGG-TMR.**

**Supplementary Figure 4.8. Protein sequences.**

Supplementary Figure 4.1

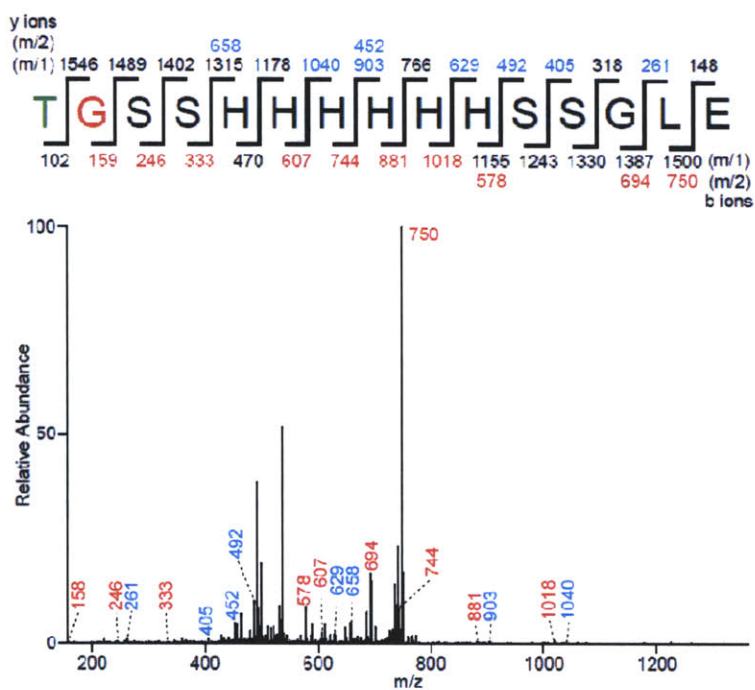


**Supporting Figure S1.** Characterization of cyclic eGFP. (a) ESI-MS of linear  $G_5$ -eGFP-LPETG-His<sub>6</sub> and circular eGFP formed by intramolecular transpeptidation. (b) A variant of eGFP lacking N-terminal glycine residues yields a mixture of unreacted starting material and hydrolysis when treated with sortase A. Conditions: eGFP-LPETG-His<sub>6</sub> (36  $\mu$ M) and sortase A (50  $\mu$ M) in (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 23 hours at 37 °C. This preparation of eGFP-LPETG-His<sub>6</sub> consisted of a mixture in which a fraction of the protein had the N-terminal Met residue cleaved during the course of protein expression to reveal an N-terminal Val residue. (c) MS/MS spectrum of a tryptic fragment of circular eGFP showing ligation of the N-terminal glycine residues (shown in red) to the C-terminal LPET residues (shown in green). Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. The full amino acid sequence of linear  $G_5$  eGFP LPETG His<sub>6</sub> is shown at the top.

Supplementary Figure 4.2

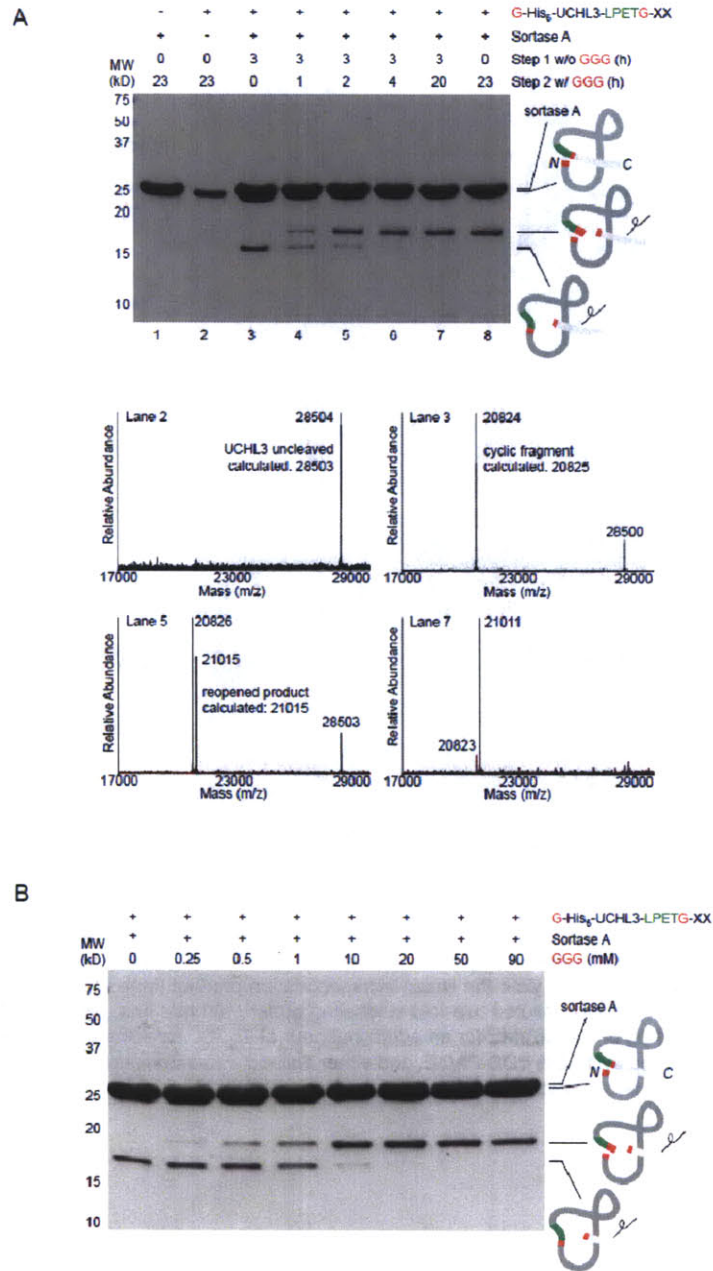
G-His<sub>5</sub>-UCHL3-LPETG-XX (amino acid sequence)

GSSHHHHHSSGLEVLFGQPHMEGQRWLPLEANPEVTNQFLKQ  
 LGLHPNWQFVDVYGMDPPELLSMVPRPVCVAVLLFPITEKEYEVFRT  
 EEEEEKISQGDVTSSVYFMKQTISNACGTIGLIHAIANNKDKMHF  
 ESGSTLKKFLEESVSMSPPEARARYLENYDAIRVTHETSAHEGQTE  
 LPETG**E**KVDLHFIALVHVDGHLVYELDGRKPFPIHGETSDETLLED  
 AIEVCKKFMERDPDELRFNAIALSAA



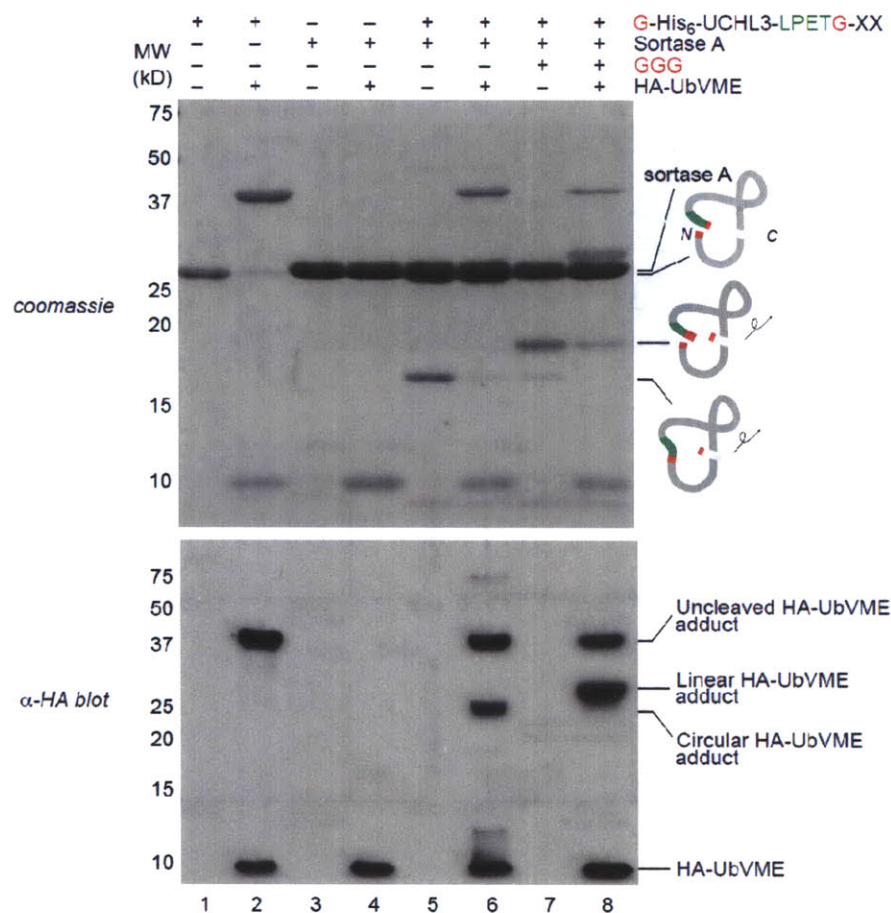


Supplementary Figure 4.3



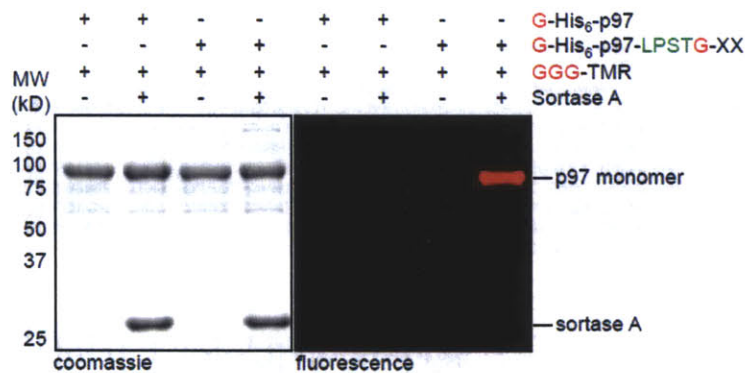
**Supporting Figure S3.** Intermolecular versus intramolecular transpeptidation for UCHL3. (a) UCHL3 cyclization is reversible. UCHL3 (30  $\mu$ M) was circularized by the addition of sortase A (150  $\mu$ M) for 3 hours. Subsequently, GGG (90 mM) was added to each reaction and incubated for the indicated times. Samples were analyzed by SDS-PAGE (top) or ESI-MS (bottom). (b) A large molar excess of GGG nucleophile is required to compete with the intramolecular UCHL3 circularization. UCHL3 (30  $\mu$ M) and sortase A (150  $\mu$ M) were incubated with the indicated concentrations of GGG nucleophile in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 3 hours at 37 °C, and analyzed by SDS-PAGE followed by coomassie staining.

Supplementary Figure 4.4



**Supporting Figure S4.** Circular UCHL3 is competent to react with a ubiquitin suicide probe. UCHL3 (30  $\mu$ M) was first circularized (lane 5) by treatment with sortase A (150  $\mu$ M) or incubated with 90 mM GGG and sortase A (150  $\mu$ M) to yield the linear transpeptidation product (lane 7) for 3 hours at 37 °C. Samples were then withdrawn, diluted two-fold in labeling buffer (100 mM Tris, pH 8.0, 150 mM NaCl) and incubated with 4  $\mu$ g of HA-UbVME for an additional hour at 25 °C. Reactions were quenched with sample buffer, analyzed by 12.5% SDS-PAGE, and either stained with coomassie (top) or transferred to nitrocellulose for anti-HA immunoblot (bottom).

**Supplementary Figure 4.5**

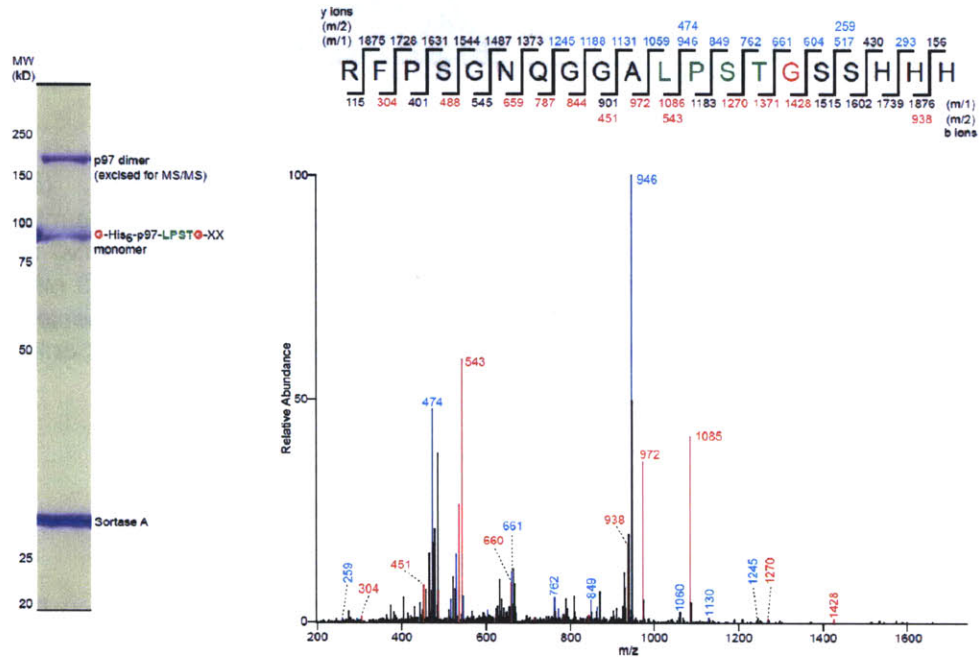


**Supporting Figure S5.** C-terminal labeling of p97. (a) G-His<sub>6</sub>-p97-LPSTG-XX (1.5 mg/mL) and G-His<sub>6</sub>-p97 (1.5 mg/mL) were incubated with sortase A (50 μM) and GGG-TMR (1 mM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 2.5 h at 37 °C. SDS-PAGE revealed the expected C-terminal transpeptidation product for G-His<sub>6</sub>-p97-LPSTG-XX only.

Supplementary Figure 4.6

G-His<sub>6</sub>-p97-LPSTG-XX (amino acid sequence)

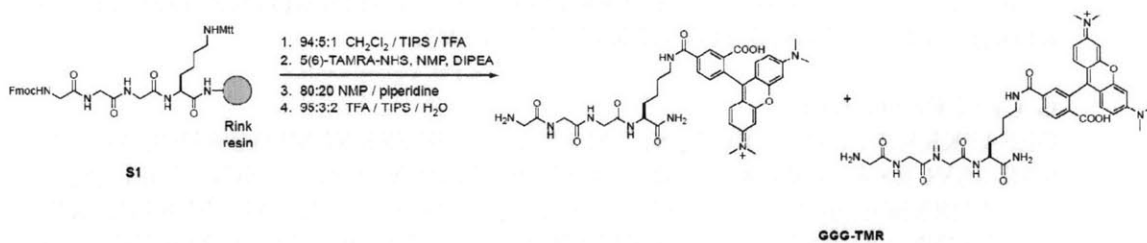
GSSHHHHHSSGLEVLVFGQPHMASGADSKGDDLSTAILKQKNRPNRLIVDEAINEDNSVVSLSQPKMDELQVFRGDTVLLKGGKRREAVCIVLS  
DDTCSDEKIRMNRVVRNLRVRLGDVISIQPCPDVKYGKRIHVLPIDDTVEGITGNLFEVYLKPYFLEAYRPIRKGDIFLVRGGMRRAVEFKVETDTP  
SPYCIAPDPTVIHCEGEPIKREDEEESLNEVGYDDIGGCRKQLAQIKEMVELPLRHPALFKAIGVKPPRGILLYGPPGTGKTLIARAVANETGAFFFL  
INGPEIMSKLAGESESNLRKAFEEAEKNAPAIIFIDELDAIAPKREKTHGEVERRIVSOLLTLMQGLKQRAHVIVMAATNRPNSIDPALRRFRGDFORE  
VDIGIPDATGRLEILQIHTKNMKLADDVDLEQVANETHGHVGDALALCSEALQAIKKMDLIDLEDETIDAEMNSLAVTMDDFRWALSQSNPS  
ALRETVVEVPQVTWEDIGGLEVDKRELQELVQYPVEHPDKFLKFGMTPSKGVLFYGPPEGKTLAKAIANECQANFISIKGPELLTMWFGSEEA  
NVREIFDKARQAAPCVLFFDELDSIAKARGGNIGDGGGAADRVINQILTEMDGMSTKKNVFIIGATNRPDIIDPAILRPGRLDQLIYIPLDEKSRVAIL  
KANLRKSPVAKDLDLEFLAKMTNGFSGADLTEICQRACKLAIRSEIEIRRETRQTNPSAMEVEEDDPVPEIRRDHFEEAMRFARRSVSDNDR  
KYEMFAQLQQSRGFGSFRFPNGQGGALPSTGSGGG



**Supporting Figure S6.** G-His<sub>6</sub>-p97-LPSTG-XX was incubated with sortase A in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 2 hours at 37 °C. The extent of oligomerization at this early time point was much less than that seen after overnight incubation. The reaction mixture was then separated by SDS-PAGE. The band corresponding to dimeric p97 was excised, digested with chymotrypsin, and the resulting peptides analyzed by MS/MS. A peptide fragment showing the ligation of the N-terminal glycine residue (shown in red) to the C-terminal LPST residues (shown in green) was identified. Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. Only the most prominent daughter ions have been labeled in the MS/MS spectrum. The full amino acid sequence of linear G-His<sub>6</sub>-p97-LPSTG-XX is shown at the top.

Supplementary Figure 4.7

Supporting Figure S7. Synthesis of GGG-TMR



**Supplementary Figure 4.8**

**Sortase A:**

MRGSHHHHHHGSKPHIDNYLHDKDKDEKIEQYDKNVKEQASKDKKQQAKPQIP  
KDKSKVAGYIEIPDADIKPEVYPGPATPEQLNRGVSF AEENASLDDQNISIAGHT  
FIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQKG  
KDKQLTLITCDDYNEKTGVWEKRKIFVATEVK

**G-Cre-LPETG-His6:**

GEFAPKKKRKVSNLLTVHQNLPALPVDATSDEVVRKNLMDMFRDRQAFSEHTW  
KMLLSVCRSWAAWCKLNNRKFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQ  
LNMLHRRSGLPRPSDSNAVSLVRRIRKENVDAGERAKQALAFERTDFDQVRS  
LMENSDRCQDIRNLAFLGIAYNLLRIAELRIRVKDISRTDGGRMLIHIGRTKTLV  
STAGVEKALSLGVTKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLS  
TRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIM  
QAGGWTNVNIVMNYIRNLDSETGAMVRLQLDGDTSGGGGSGGGGSASLPET  
GLEHHHHHHHH

**G5-eGFP-LPETG-His6:**

GGGGGRMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT  
FICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIF  
FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIM  
ADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS  
KDPNEKRDMVLEFVTAAGITLGMDEL YKLPETGRDPN SSSVDKLA AAEH  
HHHH

**eGFP-LPETG-His6**

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLT  
FICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIF  
FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIM  
ADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALS  
KDPNEKRDMVLEFVTAAGITLGMDEL YKLPETGRDPN SSSVDKLA AAEH  
HHHHHHHH

**G-His6-UCHL3-LPETG-XX:**

GSSHHHHHHSSGLEVLFGPHMEGQRWLPLEANPEVTNQFLKQLGLHPNWQ  
FVDVYGMPELLSMVPRPVCAVLLFPITEKYE VFRTEEEEKIKSQGQDVTSSV  
YFMKQTISNACGTIGLIHAIANNKDKMHFESGSLKFLFLEESVMSPEERARYLE  
NYDAIRVTHETSAHEGQTELPETGEKVDLHFIALVHVDGHL YELDGRKPFPIH  
GETSDETLLEDAIEVCKKFMERDPDELRFNAIALSAA

**G-His6-p97:**

GSSHHHHHHSSGLEVLFGPHMASGADSKGDDLSTAILKQKNRPNRLIVDEAI  
NEDNSVVSLSQPKMDELQLFRGDTVLLKGGKRREAVCIVLSDDTCSDEKIRMN  
RVVRNLRVRLGDVISIQPCPDVKYGKRIHVLPIDDTVEGITGNLFEVYLKPYFLE  
AYRPIRKGDI FLVRGGMRAVEFKVVETDPSYCVAPDTVIHCEGEPIKREDEEE  
SLNEVGYYDDIGGCRKQLAQIKEMVELPLRHPALFKAIGVKPPRGILLYGPPGTGK

Chapter 4: A straight path to circular proteins

TLIARAVANETGAFFFLINGPEIMSKLAGESSENLRKAFEEAEKNAPAIIFIDELDAI  
APKREKTHGEVERRIVSQLLTLMDGLKQRAHVIVMAATNRPNSIDPALRRFGRF  
DREVDIGIPDATGRLEILQIHTKNMKLADDVDLEQVANETHGHVVGADLAALCSE  
AALQAIRKKMDLIDLEDETIDAEVMNSLAVTMDDFRWALSQSNPSALRETVVEV  
PQVTWEDIGGLEDVKRELQELVQYPVEHPDKFLKFGMTPSKGVLFYGPPGCGKT  
LLAKAIANECQANFISIKGPELLTMWFGESEANVREIFDKARQAAPCVLFFDEL  
SIAKARGGNIGDGGGAADRVINQILTEMDGMSTKKNVFIIGATNRPDIIDPAILRP  
GRLDQLIYIPLPDEKSRVAILKANLRKSPVAKDVDLEFLAKMTNGFSGADLTEIC  
QRACKLAIRESIESEIRRERERQTNPSAMEVEEDDPVPEIRRDHFEEAMRFARR  
SVSDNDIRKYEMFAQTLQQSRGFGSFRFSPGNQGGAGPSQGSGGGTGGSVY  
TEDNDDDLYG

**G-His6-p97-LPSTG-XX:**

GSSHHHHHSSGLEVLFGPHMASGADSKGDDLSTAILKQKNRPNRLIVDEAI  
NEDNSVVSLSQPKMDELQLFRGDTVLLKGKKRREAVCIVLSDDTCSDEKIRMN  
RVVRNLRVRLGDVISIQPCPDVKYGKRIHVLPIDDTVEGITGNLFEVYLKPYFLE  
AYRPIRKGDIPLVRGGMRAVEFKVVETDPSYCVIAPDTVHCEGEPIKREDEEE  
SLNEVGYDDIGGCRKQLAQIKEMVELPLRHPALFKAIGVKPPRGILLYGPPGTGK  
TLIARAVANETGAFFFLINGPEIMSKLAGESSENLRKAFEEAEKNAPAIIFIDELDAI  
APKREKTHGEVERRIVSQLLTLMDGLKQRAHVIVMAATNRPNSIDPALRRFGRF  
DREVDIGIPDATGRLEILQIHTKNMKLADDVDLEQVANETHGHVVGADLAALCSE  
AALQAIRKKMDLIDLEDETIDAEVMNSLAVTMDDFRWALSQSNPSALRETVVEV  
PQVTWEDIGGLEDVKRELQELVQYPVEHPDKFLKFGMTPSKGVLFYGPPGCGKT  
LLAKAIANECQANFISIKGPELLTMWFGESEANVREIFDKARQAAPCVLFFDEL  
SIAKARGGNIGDGGGAADRVINQILTEMDGMSTKKNVFIIGATNRPDIIDPAILRP  
GRLDQLIYIPLPDEKSRVAILKANLRKSPVAKDVDLEFLAKMTNGFSGADLTEIC  
QRACKLAIRESIESEIRRERERQTNPSAMEVEEDDPVPEIRRDHFEEAMRFARR  
SVSDNDIRKYEMFAQTLQQSRGFGSFRFSPGNQGGALPSTGSGGG



**Supplementary Text**

**Synthesis of Triglycine Tetramethylrhodamine Peptide.** GGG-TMR was synthesized using standard Fmoc solid-phase peptide synthesis (supplemental Fig. S7). Resin bound intermediate S1 was synthesized on Rink amide resin using reported methods<sup>11</sup>. Resin S1 (72 mg, 0.036 mmol, 0.5 mmol/g resin loading) was then treated with 2 mL of 94:5:1 CH<sub>2</sub>Cl<sub>2</sub>/TIPS/TFA (5x, ~5 min each) to remove the 4-methyltrityl (Mtt) protecting group. The resin was then washed with 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x each, 3-5 min each wash) followed by 2.5 mL of NMP (3x each, 3-5 min each wash). The resin was then treated with a solution of 5(6)-TAMRA-NHS (25 mg, 0.047 mmol, mixture of 5 and 6 isomers, C1171 Invitrogen) and DIPEA (25  $\mu$ L, 0.14 mmol) in 1 mL of NMP. The reaction mixture was incubated at RT for 19 h with gentle agitation. The resin was then washed with 2.5 mL of NMP (3x each, 3-5 min each wash) followed by 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x each, 3-5 min each wash). The resin was then deprotected with 2.5 mL of 80:20 NMP/piperidine for 20 min at RT followed by washing with 2.5 mL of NMP (3x each, 3-5 min each wash) and 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x each, 3-5 min each wash). The peptide was then cleaved from the resin with 2.5 mL of 95:3:2 TFA/TIPS/H<sub>2</sub>O (5x, 15-20 min each). The combined cleavage solutions were concentrated and the crude peptide was precipitated from cold diethyl ether. The individual isomers of GGG-TMR were finally purified by RP-HPLC on a Waters Delta Pak 15  $\mu$ m, 100 Å C18 column (7.8 x 300 mm, MeCN:H<sub>2</sub>O gradient mobile phase containing 0.1% TFA, 3 mL/min). Two major peaks of roughly equal abundance were collected and gave the expected mass for GGG-TMR when analyzed by ESI-MS (LRMS, calculated for GGG-TMR [M<sup>+</sup>] = 729.3, found 729.4 for both isomers). The individual fractions were not assigned as either the 5 or 6 isomers.



*Chapter 4: A straight path to circular proteins*

Both isomers of GGG-TMR were used interchangeably in sortase-catalyzed transpeptidation reactions with no effect on protein labeling efficiency.

Supplemental Material can be found at:  
<http://www.jbc.org/content/suppl/2009/04/09/M901752200.DC1.html>THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 23, PP. 16028–16036, JUNE 5, 2009  
© 2009 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the USAA Straight Path to Circular Proteins<sup>\*[5]</sup>

Received for publication, March 16, 2009. Published, JBC Papers in Press, April 9, 2009. DOI 10.1074/jbc.M901752200

John M. Antos<sup>†1</sup>, Maximilian Wei-Lin Popp<sup>‡5</sup>, Robert Ernst<sup>‡2</sup>, Guo-Liang Chew<sup>‡3</sup>, Erik Spooner<sup>‡</sup>, and Hidde L. Ploegh<sup>‡5,4</sup>From the <sup>†</sup>Whitehead Institute for Biomedical Research and the <sup>‡</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

Folding and stability are parameters that control protein behavior. The possibility of conferring additional stability on proteins has implications for their use *in vivo* and for their structural analysis in the laboratory. Cyclic polypeptides ranging in size from 14 to 78 amino acids occur naturally and often show enhanced resistance toward denaturation and proteolysis when compared with their linear counterparts. Native chemical ligation and intein-based methods allow production of circular derivatives of larger proteins, resulting in improved stability and refolding properties. Here we show that circular proteins can be made reversibly with excellent efficiency by means of a sortase-catalyzed cyclization reaction, requiring only minimal modification of the protein to be circularized.

Sortases are bacterial enzymes that predominantly catalyze the attachment of surface proteins to the bacterial cell wall (1, 2). Other sortases polymerize pilin subunits for the construction of the covalently stabilized and covalently anchored pilus of the Gram-positive bacterium (3–5). The reaction catalyzed by sortase involves the recognition of short 5-residue sequence motifs, which are cleaved by the enzyme with the concomitant formation of an acyl enzyme intermediate between the active site cysteine of sortase and the carboxylate at the newly generated C terminus of the substrate (1, 6–8). In many bacteria, this covalent intermediate can be resolved by nucleophilic attack from the pentaglycine side chain in a peptidoglycan precursor, resulting in the formation of an amide bond between the pentaglycine side chain and the carboxylate at the cleavage site in the substrate (9, 10). In pilus construction, alternative nucleophiles such as lysine residues or diaminopimelic acid participate in the transpeptidation reaction (3, 4).

When appended near the C terminus of proteins that are not natural sortase substrates, the recognition sequence of *Staphylococcus aureus* sortase A (LPXTG) can be used to effectuate a sortase-catalyzed transpeptidation reaction using a diverse array of artificial glycine-based nucleophiles (Fig. 1). The result is efficient installation of a diverse set of moieties, including

lipids (11), carbohydrates (12), peptide nucleic acids (13), biotin (14), fluorophores (14, 15), polymers (16), solid supports (16–18), or peptides (15, 19) at the C terminus of the protein substrate. During the course of our studies to further expand sortase-based protein engineering, we were struck by the frequency and relative ease with which intramolecular transpeptidation reactions were occurring. Specifically, proteins equipped with not only the LPXTG motif but also N-terminal glycine residues yielded covalently closed circular polypeptides (Fig. 1). Similar reactivity using sortase has been described in two previous cases; however, rigorous characterization of the circular polypeptides was absent (16, 20). The circular proteins in these reports were observed as minor components of more complex reaction mixtures, and the cyclization reaction itself was not optimized.

Here we describe our efforts toward applying sortase-catalyzed transpeptidation to the synthesis of circular and oligomeric proteins. This method has general applicability, as illustrated by successful intramolecular reactions with three structurally unrelated proteins. In addition to circularization of individual protein units, the multiprotein complex AAA-ATPase p97/VCP/CDC48, with six identical subunits containing the LPXTG motif and an N-terminal glycine, was found to preferentially react in daisy chain fashion to yield linear protein fusions. The reaction exploited here shows remarkable similarities to the mechanisms proposed for circularization of cyclotides, small circular proteins that have been isolated from plants (21–23).

## EXPERIMENTAL PROCEDURES

**Synthesis of Triglycine Tetramethylrhodamine Peptide**—The structure of GGG-TMR<sup>5</sup> and a detailed synthetic protocol are provided in the supplemental material and supplemental Fig. S7.

**Cloning and Protein Expression**—Full amino acid sequences for all proteins used in this study are given in supplemental Fig. S8.

Recombinant sortase A (residues 26–206) containing an N-terminal hexahistidine tag was produced in *Escherichia coli* as described previously (8). Purified sortase A was stored in 10% glycerol, 50 mM Tris, pH 8.0, 150 NaCl at –80 °C until further use.

\* This work was supported, in whole or in part, by National Institutes of Health Grant R21-EB008875.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental text and reference and Figs. S1–S8.

<sup>1</sup> Recipient of a Clay postdoctoral fellowship.

<sup>2</sup> Recipient of EMBO Long Term Fellowship ALTF 379 2008.

<sup>3</sup> Supported by Undergraduate Research Opportunities Program from Massachusetts Institute of Technology.

<sup>4</sup> To whom correspondence should be addressed. E-mail: ploegh@wi.mit.edu.

<sup>5</sup> The abbreviations used are: GGG-TMR, triglycine tetramethylrhodamine peptide; eGFP, enhanced green fluorescent protein; UCHL3, ubiquitin C-terminal hydrolase L3; GGG, triglycine peptide; ESI-MS, electrospray ionization mass spectrometry; RP-HPLC, reversed-phase high performance liquid chromatography; MS/MS, tandem mass spectrometry; PDB, Protein Data Bank; Ni-NTA, nickel-nitrilotriacetic acid.

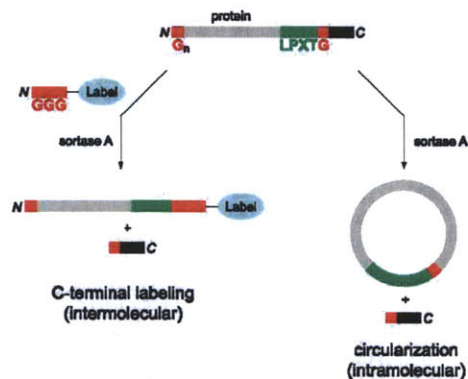


FIGURE 1. Protein substrates equipped with a sortase A recognition sequence (LPXTG) can participate in intermolecular transpeptidation with synthetic oligoglycine nucleophiles (left) or intramolecular transpeptidation if an N-terminal glycine residue is present (right).

G-Cre-LPETG-His<sub>6</sub> was cloned into the pTriEx-1.1 Neo expression vector (Novagen) using standard molecular biology techniques. The construct contains two point mutations (M117V and E340Q) and a flexible spacer (GGGGSGGGGS) inserted before the LPETG sortase recognition site. G-Cre-LPETG-His<sub>6</sub> was expressed and purified using procedures similar to those reported previously for HTNCre (24). G-Cre-LPETG-His<sub>6</sub> was first transformed into Tuner (DE3) pLacI cells (Novagen), and a starter culture was grown in sterile LB media supplemented with 1% (w/v) glucose, chloramphenicol (34  $\mu\text{g}/\text{ml}$ ), and ampicillin (100  $\mu\text{g}/\text{ml}$ ). This culture was used to inoculate a large scale culture of sterile LB containing chloramphenicol (34  $\mu\text{g}/\text{ml}$ ) and ampicillin (100  $\mu\text{g}/\text{ml}$ ). G-Cre-LPETG-His<sub>6</sub> was expressed after a 3-h induction with isopropyl 1-thio- $\beta$ -D-galactopyranoside (0.5 mM) at 37 °C. Cells were resuspended in 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 20 mM imidazole, pH 8.0. The suspension was adjusted to 50  $\mu\text{g}/\text{ml}$  DNase I, 460  $\mu\text{g}/\text{ml}$  lysozyme, and 1 mM MgCl<sub>2</sub> and incubated at 4 °C for 1.5 h. The suspension was then sonicated and centrifuged. The clarified lysate was then treated with Ni-NTA-agarose (Qiagen) for 1 h at 4 °C. The resin was washed with 12 column volumes of 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 20 mM imidazole, pH 8.0, followed by 4 column volumes of 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 30 mM imidazole, pH 8.0. The protein was then eluted with 10 mM Tris, 100 mM phosphate, 300 mM NaCl, and 300 mM imidazole, pH 8.0. The purified protein was then dialyzed first against 20 mM Tris, pH 7.5, 500 mM NaCl followed by 50% glycerol, 20 mM Tris, pH 7.5, 500 mM NaCl. G-Cre-LPETG-His<sub>6</sub> was then passed through a 0.22- $\mu\text{m}$  filter to remove minor precipitation and stored at 4 °C.

G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> was prepared from a previously reported eGFP construct lacking the five N-terminal glycine residues using a QuickChange® II site-directed mutagenesis kit (Stratagene) and produced in *E. coli* using reported procedures (11). Purified G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> was buffer exchanged into 20 mM Tris, pH 8.0, 150 mM NaCl and stored at 4 °C. UCHL3 with the sortase recognition sequence (LPETG) substituted for

amino acids 159–163 was cloned and produced in *E. coli* as described previously (25).

Human p97 (806 amino acids) was PCR-amplified and cloned via the NdeI and HindIII restriction sites into a pET28a+ expression vector (Novagen) to yield the G-His<sub>6</sub>-p97 construct. G-His<sub>6</sub>-p97-LPSTG-XX was generated by introducing two point mutations (G782L and Q785T) and a stop codon at position 791 using QuickChange® mutagenesis (Stratagene). Recombinant p97 was expressed at 30 °C in *E. coli* after induction for 3 h with 0.5 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside. Cells were resuspended in buffer A (50 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol, 20 mM imidazole, and 7.1 mM  $\beta$ -mercaptoethanol), adjusted to 15  $\mu\text{g}/\text{ml}$  lysozyme and 10  $\mu\text{g}/\text{ml}$  DNase I, and lysed by two passes through a French pressure cell at 1200 p.s.i. After centrifugation for 30 min at 40,000  $\times g$ , the supernatant was bound to nickel-Sepharose resin (GE Healthcare). After washing the resin with 20 column volumes of buffer A, p97 was eluted with buffer A containing 250 mM imidazole. Hexameric rings of p97 were further purified on a Superdex 200 HR 16/60 column (GE Healthcare) using 25 mM Tris, pH 8.0, 150 mM KCl, 2.5 mM MgCl<sub>2</sub>, 5% glycerol as the mobile phase. The purified protein was snap-frozen and stored at -80 °C.

**Circularization and Intermolecular Transpeptidation**—Transpeptidation reactions were performed by combining the necessary proteins/reagents at the specified concentrations in the presence of sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) and incubating at 37 °C for the times indicated. Diglycine and triglycine (GGG) peptides were purchased from Sigma. Reactions were halted by the addition of reducing Laemmli sample buffer and analyzed by SDS-PAGE. Gels were visualized by staining with Coomassie Blue. Fluorescence was visualized on a Typhoon 9200 Imager (GE Healthcare). Crude reactions were also diluted into either 0.1% formic acid or water for ESI-MS analysis. ESI-MS was performed on a Micromass LCT mass spectrometer (Micromass® MS Technologies) and a Paradigm MG4 HPLC system equipped with an HTC PAL autosampler (Michrom BioResources) and a Waters symmetry 5- $\mu\text{m}$  C8 column (2.1  $\times$  50 mm, MeCN:H<sub>2</sub>O (0.1% formic acid) gradient mobile phase, 150  $\mu\text{l}/\text{min}$ ).

**Purification and Refolding of eGFP**—G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> (50  $\mu\text{M}$ ) was circularized by treatment with sortase A (50  $\mu\text{M}$ ) in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 24 h at 37 °C. The reaction was run on a 750- $\mu\text{l}$  scale. The entire reaction was then diluted into 10 ml of 20 mM Tris, 500 mM NaCl, and 20 mM imidazole, pH 8.0. This solution was then applied to a column consisting of 2 ml of Ni-NTA-agarose (Qiagen) pre-equilibrated with 20 mM Tris, 500 mM NaCl, and 20 mM imidazole, pH 8.0. The flow-through was then concentrated and buffer exchanged in 20 mM Tris, 150 mM NaCl, pH 8.0, using a NAP™ 5 Sephadex™ column (GE Healthcare). The concentrations of circular eGFP and linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> were estimated by UV-visible spectroscopy using the absorbance of eGFP at 488 nm (extinction coefficient 55,900 M<sup>-1</sup> cm<sup>-1</sup>) (26). Circular and linear eGFP (40  $\mu\text{l}$  of 18  $\mu\text{M}$  solutions) was placed in 1.5-ml microcentrifuge tubes and denatured by heating to 90 °C for 5 min. Samples were then incubated at room temperature in the dark for the times indi-

## Circular Proteins

cated. Fluorescent images were acquired using a UV gel documentation system (UVP Laboratory Products).

**Reaction of Cyclic UCHL3 with Activity-based Ubiquitin Probe**—UCHL3 (30  $\mu\text{M}$ ) was incubated with sortase A (150  $\mu\text{M}$ ) in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM  $\text{CaCl}_2$ ) in the presence or absence of 90 mM GGG peptide (Sigma) on a 25- $\mu\text{l}$  scale at 37  $^\circ\text{C}$  for 3 h. Ten microliters was withdrawn and diluted with 10  $\mu\text{l}$  of labeling buffer (100 mM Tris, pH 7.5, 150 mM NaCl). Hemagglutinin epitope-tagged ubiquitin vinyl methyl ester (4  $\mu\text{g}$ ) was added as well as 1 mM dithiothreitol and incubated at room temperature for 1 h. Reactions were then separated on an SDS-polyacrylamide gel and visualized by Coomassie staining or  $\alpha$ -HA immunoblot (supplemental Fig. S4). Hemagglutinin epitope-tagged ubiquitin vinyl methyl ester was prepared following published protocols (27).

**MS/MS Sequencing of Proteolytic Fragments from Circular Proteins**—Prior to MS/MS analysis, circular eGFP and Cre were separated from sortase A by RP-HPLC using an Agilent 1100 Series HPLC system equipped with a Waters Delta Pak 5  $\mu\text{m}$ , 100  $\text{\AA}$  C18 column (3.9  $\times$  150 mm, MeCN:H<sub>2</sub>O gradient mobile phase containing 0.1% trifluoroacetic acid, 1 ml/min). Fractions containing the circular proteins were pooled and subjected to trypsin digestion. Crude transpeptidation reactions containing circular UCHL3 were separated by SDS-PAGE followed by Coomassie staining. The band corresponding to circular UCHL3 was excised and digested with Glu-C. Crude transpeptidation reactions containing dimeric p97 were separated by SDS-PAGE followed by Coomassie staining. The transpeptidation reaction used for this purpose was incubated for only 2 h and therefore contains less oligomerization than that seen after an overnight incubation (see supplemental Fig. S6). The band corresponding to dimeric p97 was excised and digested with chymotrypsin. For all protein substrates, the peptides generated from proteolytic digestion were extracted and concentrated for analysis by RP-HPLC and tandem mass spectrometry. RP-HPLC was carried out on a Waters NanoAcquity HPLC system with a flow rate of 250 nl/min and mobile phases of 0.1% formic acid in water and 0.1% formic acid in acetonitrile. The gradient used was isocratic 1% acetonitrile for 1 min followed by 2% acetonitrile per min to 40% acetonitrile. The analytical column was 0.075  $\mu\text{m}$   $\times$  10 cm with the tip pulled to 0.005  $\mu\text{m}$  and self-packed with 3  $\mu\text{m}$  Jupiter C18 (Phenomenex). The column was interfaced to a Thermo LTQ linear ion trap mass spectrometer in a nanospray configuration, and data were collected in full scan mode followed by MS/MS analysis in a data-dependent manner. The mass spectral data were data base searched using SEQUEST.

**Construction of Molecular Models**—Molecular models were generated from published crystal structures (PDB codes 1kbu, 1gfl, 1xd3, and 3cfl) (28–31). N- and C-terminal residues were added using Coot 0.5 (32). Protein termini were repositioned using the Auto Sculpting function in MacPyMOL (DeLano Scientific LLC). Residues visible in the published crystal structures were not moved during positioning of the extended N and C termini. All protein images in this study were generated using MacPyMOL.

## RESULTS

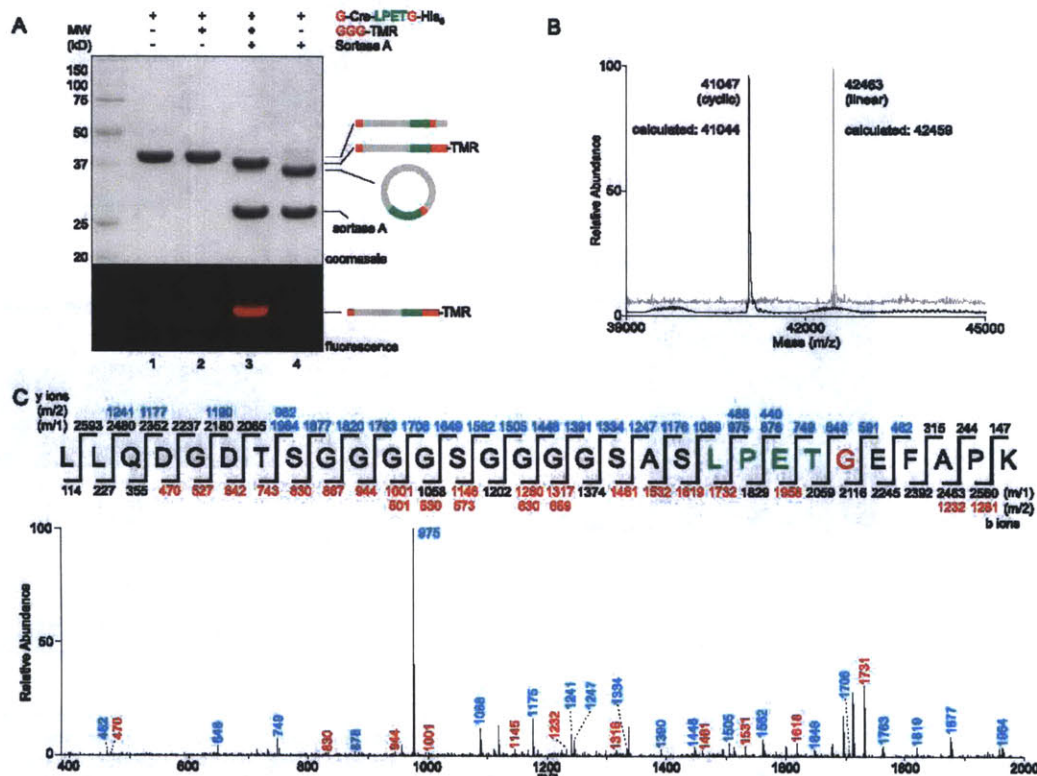
**Cre Recombinase**—We first noticed the presence of a circular protein product when installing a C-terminal modification onto a nonfunctional mutant of Cre recombinase containing a single N-terminal glycine residue and the requisite LPETG sequence near the C terminus. The LPETG motif was separated from the native protein by a flexible amino acid linker (GGGGSGGGGS). Whereas installation of the label at the Cre C terminus proceeded efficiently when a triglycine nucleophile containing tetramethylrhodamine (GGG-TMR) was included, we observed a product that migrated more rapidly on SDS-PAGE when nucleophile was omitted from the reaction mixture (Fig. 2A). Hydrolysis of the sortase acyl enzyme is known to proceed slowly in the absence of glycine nucleophiles (19, 33, 34). However, when reaction mixtures were analyzed by ESI-MS, we consistently observed a protein species that differed from the mass expected for hydrolysis by approximately –18 Da (Fig. 2B). This mass was consistent with intramolecular nucleophilic attack, suggesting that the single N-terminal glycine residue was serving as the nucleophile in this transformation. Ultimately, MS/MS on tryptic digests of this species showed unequivocally that it consisted of a covalently closed circular product of Cre, with the N-terminal glycine fused exactly at the LPETG cleavage site in the expected position (Fig. 2C).

Recognizing that the LPETG motif is maintained in the cyclized Cre product, we suspected that sortase should be capable of cleaving the circular protein at this site, thus producing an equilibrium between circular and linear forms of Cre. To demonstrate this point, Cre was first incubated with sortase in the presence or absence of triglycine nucleophile (Fig. 3A). A portion of the cyclized reaction mixture (Fig. 3A, lane 1) was then treated with a large molar excess of triglycine nucleophile or left alone for a further 24 h (Fig. 3A, lanes 2 and 3). Remarkably, upon treatment with exogenous nucleophile, the pre-cyclized material yielded a reaction mixture that was nearly identical to the result obtained when nucleophile was included from the very beginning of the experiment (Fig. 3A, compare lanes 3 and 4). This result provided further evidence that cyclized Cre indeed contains the expected LPETG motif at the site of covalent closure. In addition, it suggested that hydrolysis of the acyl enzyme intermediate does not effectively compete during cyclization, because the hydrolyzed material should be unable to participate in the transpeptidation reaction.

The circularization reaction observed for Cre proceeded with remarkable efficiency. Conversion was estimated to be >90% by SDS-PAGE. By taking an existing crystal structure (29) of the Cre protein and modeling in those residues not visible in the structure, it was clear that the N and C termini were located in sufficiently close proximity to permit closure without significant perturbation of the native structure (Fig. 3B). We assume that these regions possess considerable flexibility because they are not resolved in the crystal structure.

**eGFP**—Having verified the cyclization of Cre recombinase, we sought to explore the generality of this technique. To this end we generated a derivative of eGFP containing the LPETG sequence and five N-terminal glycine residues. This construct was of particular interest because inspection of the x-ray crystal





**FIGURE 2. Cyclization of Cre recombinase.** A, G-Cre-LPETG-His<sub>6</sub> (50  $\mu$ M) was incubated with sortase A (50  $\mu$ M) in the presence or absence of fluorescent GGG-TMR (10 mM) in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 21.5 h at 37 °C. SDS-PAGE revealed the expected C-terminal transpeptidation product when GGG-TMR was included, whereas omission of the triglycine nucleophile resulted in clean conversion to a unique protein species with a lower apparent molecular weight. Schematic representations to the right of the gel indicate the topology of the protein species produced by transpeptidation. B, ESI-MS of linear G-Cre-LPETG-His<sub>6</sub> and circular Cre formed by intramolecular transpeptidation. C, MS/MS spectrum of a tryptic fragment of circular Cre showing the ligation of the N-terminal residues (GEFAPK) to the C-terminal LPET motif. Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. Only the most prominent daughter ions have been labeled in the MS/MS spectrum.

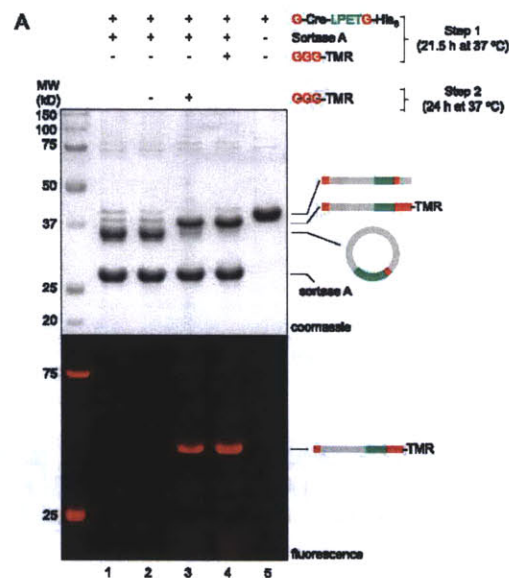
structure (31) revealed that the N and C termini were positioned on the same end of the  $\beta$ -barrel, suggesting that this substrate should be ideal for cyclization (Fig. 4A). Furthermore, in one of the earliest reports on the use of sortase for protein engineering, a similar eGFP substrate was described and reported to cyclize in the presence of sortase (16). In this instance, cyclization only proceeded in modest yield, and the putative cyclized product was produced as a mixture with higher molecular weight species assigned as oligomers of eGFP formed by intermolecular transpeptidation. Thus, to explore potential complications caused by intermolecular reactions, we studied the reaction of our eGFP construct in the presence of sortase.

In our hands, we observed clean conversion to a lower molecular weight species (>90% estimated conversion) with little to no evidence for oligomerization (Fig. 4B). A higher molecular weight polypeptide was observed at early time points and may

represent a covalent eGFP dimer that is generated transiently over the course of the reaction. Higher molecular species, however, were only observed in trace quantities in the final reaction mixture. As in the case of Cre, evidence for circularization was provided by mass spectral characterization of the intact circular protein and MS/MS sequencing of tryptic peptides (supplemental Fig. S1). As an additional control to demonstrate that the N-terminal glycine residue was the only nucleophile participating in intramolecular transpeptidation, we analyzed the behavior of an eGFP derivative that lacked an N-terminal glycine. In this case, ESI-MS revealed products consistent with hydrolysis of the acyl enzyme intermediate, rather than intramolecular nucleophilic attack (supplemental Fig. S1).

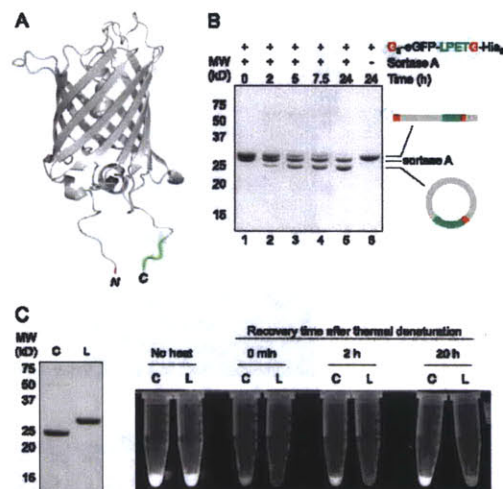
Circularization has been shown to confer unique properties onto proteins when compared with the linear form (35–37). In the case of GFP circularized using intein-based methods, these properties include a reduced rate of unfolding when exposed to

## Circular Proteins



**FIGURE 3. Cyclization of Cre is reversible.** A G-Cre-LPETG-His<sub>6</sub> (50 μM) was circularized by treatment with sortase A (50 μM) in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 21.5 h at 37 °C (lane 1). This reaction mixture was then treated with 10 mM GGG-TMR (lane 2) or simply incubated for an additional 24 h at 37 °C (lane 3). All reactions were analyzed by SDS-PAGE with visualization by Coomassie staining or in-gel fluorescence. For comparison, a C-terminal labeling reaction performed using 10 mM GGG-TMR without prior cyclization of the Cre substrate (lane 4) and a sample of linear G-Cre-LPETG-His<sub>6</sub> incubated without sortase A or nucleophile (lane 5) are included. **B**, molecular model of Cre recombinase monomer (generated from PDB code 1kbu) (29) showing the proximity relationship between the N and C termini. The N-terminal glycine residue is highlighted in red, and the C-terminal LPET residues are shown in green.

denaturants, as well as an enhanced rate of refolding following denaturation (35). We observed a similar phenomenon for eGFP circularized using sortase (Fig. 4C). Circular eGFP was first separated from residual sortase A using Ni-NTA resin. This material retained fluorescence suggesting that covalent



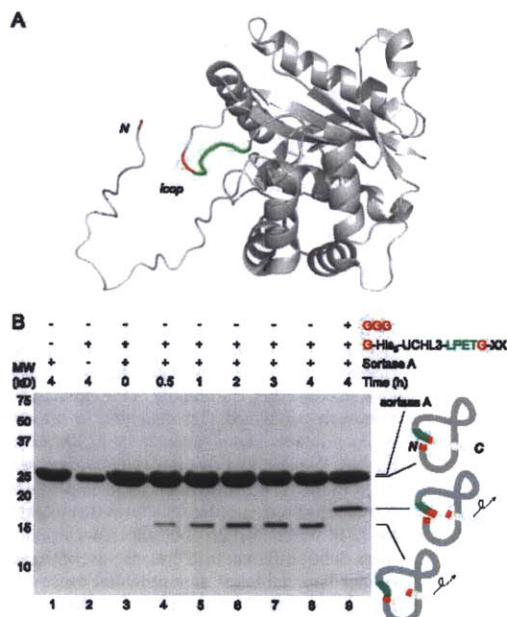
**FIGURE 4. Circularization of eGFP.** A, molecular model of eGFP (generated from PDB code 1gfl) (31) showing the proximity relationship between the N and C termini. The N-terminal glycine residue is highlighted in red, and the C-terminal LPET residues are shown in green. B, G<sub>2</sub>-eGFP-LPETG-His<sub>6</sub> (50 μM) was circularized by treatment with sortase A (50 μM) in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 24 h at 37 °C. Schematic representations to the right of the gel indicate the topology of the protein species produced by transpeptidation. C, circular eGFP (C) recovers fluorescence more rapidly than linear G<sub>2</sub>-eGFP-LPETG-His<sub>6</sub> (L) following thermal denaturation for 5 min at 90 °C. SDS-PAGE analysis confirmed the purity and concentration of the circular eGFP (C) and linear G<sub>2</sub>-eGFP-LPETG-His<sub>6</sub> (L) samples used for refolding.

ligation of the N and C termini had minimal impact on the structure of this substrate. Circular and linear eGFP were then subjected to simple thermal denaturation, followed by recovery at room temperature. As shown in Fig. 4C, circular eGFP regained fluorescence more rapidly than linear eGFP.

**UCHL3**—Even an internally positioned LPXTG motif was sufficient to effectuate a circularization reaction. We installed a sortase recognition site in the crossover loop of the ubiquitin C-terminal hydrolase UCHL3, and we demonstrated that the continuity of the polypeptide backbone can be disrupted with concomitant installation of a covalent modification that reports on the accuracy of cleavage and transpeptidation (25). This reaction proceeds without complete loss of activity of UCHL3, indicating that even the cleaved form of UCHL3 retains its structural integrity to a significant degree (25). This UCHL3 construct was prepared with an N-terminal glycine residue, and examination of the crystal structure of UCHL3 (30) clearly showed the close apposition of the N terminus and the crossover loop, suggesting that cyclization to yield a circular fragment containing the N-terminal portion of UCHL3 should be readily observable (Fig. 5A).

As expected, in the absence of added nucleophile, the N-terminal glycine serves as a highly efficient nucleophile to yield a circular fragment that contains the N-terminal portion of UCHL3 (Fig. 5B). The identity of the circular polypeptide was confirmed by MS/MS of the peptide containing the expected fusion of the N-terminal glycine residue with the new C termi-





**FIGURE 5. UCHL3 with an internally positioned LPETG motif is circularized by sortase A.** A, molecular model of human UCHL3 (generated from PDB code 1xd3) (30) showing the active site crossover loop bearing an LPETG substitution (LPET residues shown in green and Gly residue shown in red). The N-terminal glycine residue that serves as the nucleophile for intramolecular transpeptidation is highlighted in red. B, UCHL3 (30  $\mu$ M) bearing an LPETG substitution in the active site crossover loop was incubated with sortase A (150  $\mu$ M) in the absence or presence of GGG nucleophile (90 mM) in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM  $\text{CaCl}_2$ ) for the indicated times at 37°C and analyzed by SDS-PAGE, followed by Coomassie staining. Schematic representations to the right of the gel indicate the topology of the protein species produced by transpeptidation.

nus released from the crossover loop (see supplemental Fig. S2). Cyclization was efficiently blocked if a high concentration of triglycine (GGG) was included in the reaction, generating instead the N-terminal fragment of UCHL3 transacylated onto the triglycine nucleophile (Fig. 5B, lane 9, and supplemental Fig. S3). Cyclization could also be reversed by adding an excess of triglycine to reaction mixtures preincubated with sortase to allow cyclization. This reopening reaction was observed by both SDS-PAGE and ESI-MS (supplemental Fig. S3).

To test the functional properties of cyclic UCHL3, we incubated reaction mixtures with an activity-based probe consisting of ubiquitin equipped with an electrophilic vinyl methyl ester moiety at the C terminus (supplemental Fig. S4). Probes of this nature are able to specifically alkylate active site cysteine residues in ubiquitin-specific hydrolases such as UCHL3 (25, 27, 38). Following circularization, the active site cysteine (Cys-95) of UCHL3 is located in the circular N-terminal fragment, and indeed we observed covalent labeling of this fragment with a corresponding shift in apparent molecular weight consistent with the attachment of ubiquitin. This result suggests that despite cleavage of the polypeptide backbone, the circular

N-terminal fragment of UCHL3 and the C-terminal portion released during transpeptidation remain associated and preserve the affinity of UCHL3 for ubiquitin. This result is consistent with previous observations from our laboratory demonstrating that covalent closure of the UCHL3 crossover loop is dispensable for enzyme activity (25).

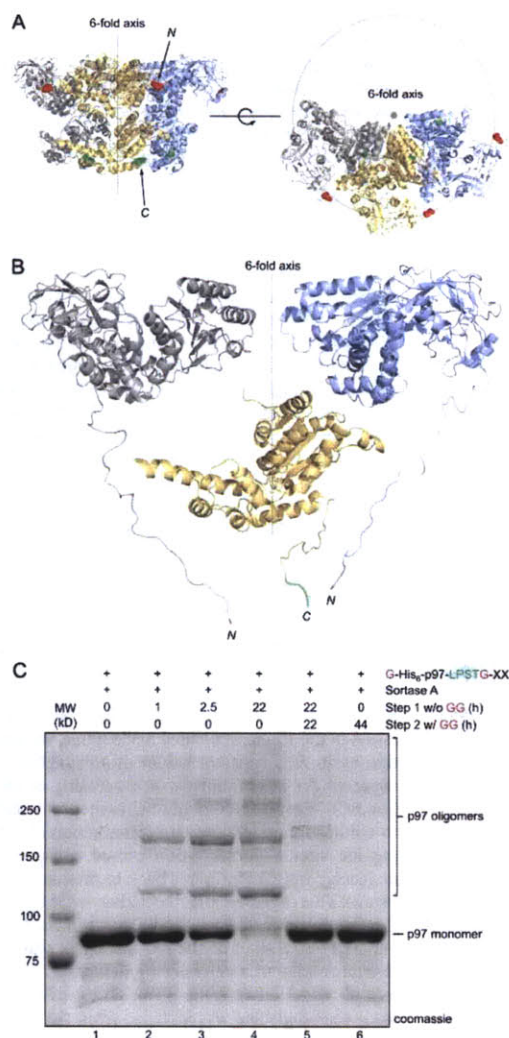
**p97**—The above examples concern single chain proteins whose termini are sufficiently close to allow covalent closure by means of the sortase-mediated transacylation reaction. Similar proximity relationships between protein termini should also be present on separate polypeptides that assemble into defined oligomeric structures. As an example, we examined p97, a hexameric AAA-ATPase. We generated a derivative of p97 (G-His<sub>6</sub>-p97-LPSTG-XX) containing an LPSTG motif near the C terminus, and a hexahistidine tag capped by two serine residues and a single glycine at the N terminus. The structure of a p97 trimer in the presence of ADP has been solved at 3.5 Å resolution (28), with several residues from the N and C termini not visible (Fig. 6A). When all the residues present in our modified version of p97 were modeled onto the published trimer of p97, it was evident that the N and C termini of adjacent p97 units were sufficiently close to permit covalent cross-linking (Fig. 6B). G-His<sub>6</sub>-p97-LPSTG-XX was expressed in *E. coli* and yielded the hexameric p97 ring, as assessed by gel filtration. As expected, this derivative of p97 was an excellent substrate for transpeptidation at its C terminus, allowing efficient installation of a label when incubated in the presence of sortase and GGG-TMR (supplemental Fig. S5). In contrast, a variant of p97 lacking the LPSTG sequence showed no labeling (supplemental Fig. S5). When G-His<sub>6</sub>-p97-LPSTG-XX was treated with sortase A in the absence of added nucleophile, we observed formation of an SDS-resistant ladder of polypeptides, as would be expected for intermolecular cross-linking of p97 monomers (Fig. 6C). We were confident that these species arise from head-to-tail ligation of p97 because introduction of excess diglycine peptide after oligomerization caused collapse of the higher molecular weight structures back to monomeric p97 (Fig. 6C, lane 5). This suggested that the higher order aggregates are held together by newly formed LPSTG units formed from the C-terminal LPST residues of one p97 monomer and the N-terminal glycine residue of a neighboring monomer. The banding pattern observed for reopening was also nearly identical to that seen when diglycine was included from the very beginning of the experiment, a scenario where installation of diglycine at the C terminus of each p97 subunit is presumed to be the major reaction pathway (Fig. 6C, lane 6). We have also been able to identify peptides consistent with intermolecular cross-linking of p97 subunits by MS/MS (supplemental Fig. S6).

## DISCUSSION

Cyclic proteins are an interesting class of polypeptides that often display unique properties because of covalent closure of the amide backbone (39, 40). Although some cyclic protein derivatives occur naturally, methods for generating cyclic proteins in the laboratory provide a means for accessing cyclic versions of proteins that only occur in linear form. Intramolecular sortase-catalyzed transpeptidation provides a straightforward method for accessing these types of cyclic proteins. The



## Circular Proteins



**FIGURE 6. Positions of the N and C termini in the p97 hexamer are suitable for intermolecular cross-linking through sortase-catalyzed transpeptidation.** A, molecular model of p97 trimer (generated from PDB code 3cf1) (28) showing the relative position of p97 monomers in the hexameric ring. The visible N and C termini from the published p97 trimer structure are indicated in red and green, respectively. B, molecular model of G-His<sub>6</sub>-p97-LPSTG-XX showing the proximity relationship between N and C termini in adjacent p97 monomers. N- and C-terminal residues not visible in the published crystal structure have been modeled onto the existing structure. N-terminal glycine residues are shown in red, and the C-terminal LPST residues are shown in green. For clarity, the C-terminal domains of the outer monomers are hidden, as is the N-terminal domain of the central monomer. C, G-His<sub>6</sub>-p97-LPSTG-XX (1.5 mg/ml) was incubated with sortase A (30 μM) in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for the times indicated at 37 °C. After 22 h, diglycine (GG) was added (100 mM final concentration), resulting in disappearance of the covalent oligomers (lane 5). For comparison, a control reaction containing 100 mM diglycine peptide from the outset of the experiment is shown in lane 6.

transpeptidation reaction described here bears a remarkable resemblance to the proposed biosynthesis of the largest class of naturally occurring cyclic proteins, the cyclotides (21–23). In both cases, linear protein precursors are cleaved by cysteine proteases to generate an acyl-enzyme intermediate that is subsequently resolved by nucleophilic attack from the N terminus of the linear proteins to generate the cyclic product.

In this work we have explored transpeptidation reactions using four structurally diverse protein substrates. Cyclization has been confirmed for three proteins, including an example (UCHL3) utilizing an LPXTG sequence positioned in a flexible internal loop rather than near the C terminus of the protein. Cyclization and oligomerization via sortase-mediated transpeptidation have been previously suggested to occur for an eGFP construct modified in a manner similar to that used here (16), and for a by-product from a protein purification system where the substrate circularized appears to be sortase A itself (20). In both cases, the identity of the circular products was not rigorously confirmed. Our data identify the circular or oligomeric products unambiguously by MS/MS for all substrates studied. We also find that our eGFP derivative strongly favors cyclization over oligomerization, showing little evidence for the formation of higher order structures that might be expected by the head-to-tail ligation of termini from separate eGFP monomers. Subtle differences in the structure of the eGFP constructs cannot be overlooked as a potential cause for the observed results. For example, our eGFP is extended at the N terminus by only five glycine residues, whereas the construct studied by Parthasarathy *et al.* (16) contains an additional 17 residues, including 3 N-terminal glycines. Future work will be required to thoroughly characterize the effect of distance relationships between protein termini on favoring intra- versus intermolecular transpeptidation.

With respect to protein cyclization, sortase-mediated circularization is efficient despite the potential for competing reaction pathways. In the absence of added oligoglycine nucleophile, these include hydrolysis of the acyl enzyme intermediate, reattachment of the C-terminal protein fragment that is lost upon initial cleavage of the protein substrate by sortase, or, as mentioned above, oligomerization of protein monomers in head-to-tail fashion. Even when oligoglycine nucleophile is added with the intent of blocking the cyclization pathway, millimolar concentrations are necessary to efficiently compete with cyclization. One factor that certainly must contribute to this observed preference for cyclization is the distance between protein termini. Inspection of the data base of PDB shows that nearly one-third of proteins with known structures have their termini in rather close apposition (within 20 Å) (40). The LPXTG sequence itself spans roughly 15 Å in an extended conformation, suggesting that circularization via sortase-catalyzed transpeptidation might be amenable to a significant fraction of proteins using the LPXTG sequence alone to bridge the gap between N and C termini. Larger distances could simply be covered by inserting flexible amino acid spacers at either terminus. We also consider it likely that the circularized version of a protein will show more restricted mobility in the segment that corresponds to the newly established LPXTG connection between its termini. This fact alone may render the circular



product a comparatively worse substrate for sortase and therefore assist in driving the transpeptidation reaction toward cyclization. As evidence for this point, we have observed previously that sortase fails to cleave LPXTG motifs placed in structured loops of class I major histocompatibility complex molecules (14).

Sortase-catalyzed transpeptidation provides an attractive alternative to existing methods for peptide and protein circularization. Chemical synthesis can provide access to circular polypeptides of modest size, with circularization of linear precursors having been achieved using native chemical ligation (41–43), subtiligase (44), or standard amide bond-forming reactions common to solid-phase peptide synthesis (43, 45). For larger proteins beyond the technical capabilities of solid-phase synthesis, cyclization is most often accomplished using native chemical ligation, typically in conjunction with split-intein expression constructs (35–37, 46–48). When compared with the split-intein approach, the modest modification necessary to render proteins amenable to cyclization or oligomerization is certainly an attractive feature of the sortase-catalyzed process. Proteins must simply possess a sortase recognition sequence (LPXTG) either near the C terminus or in a flexible loop and an N-terminal glycine residue to act as the nucleophile. These modifications are not anticipated to have a significant impact on protein expression or function. In contrast, protein circularization by split-intein methods requires more extensive modifications of the expression construct, a necessity that may reduce protein expression or affect protein solubility. It should be noted, however, that the number of extra amino acid residues at the site of N- to C-terminal ligation following excision of the large intein domains can be less than the five residues (LPXTG) that remain after circularization using sortase A.

The sortase-catalyzed approach also provides additional levels of control over the ensuing transpeptidation reaction. This may be particularly useful for oligomeric species, such as the p97 example described here. Specifically, our modified p97 protein (G-His<sub>6</sub>-p97-LPSTG-XX) is produced in a form that is by itself unreactive. This allows protein expression and the subsequent assembly and purification of the hexamer to be completed first, without complications caused by premature covalent oligomerization. Cross-linking is then induced by the addition of sortase after the individual subunits have been correctly positioned in the hexameric ring. The extent of transpeptidation can be further controlled by inclusion of synthetic oligoglycine nucleophiles, either during the transpeptidation reaction or after transpeptidation is complete. The latter scenario even allows cyclization to be completely reversed. Incubating circular protein products with sortase in the presence of an oligoglycine nucleophile restores linearity to the protein product, because in the course of the initial cyclization reaction, the LPXTG motif is restored. An equilibrium between closed and open forms is thus established and can be driven toward the linear state by adding a large excess of the oligoglycine nucleophile.

The implications of protein cyclization or oligomerization for protein engineering are numerous. In the case of protein oligomerization, the ability to link protein subunits held in a defined geometry might be exploited to explore subtle changes

in intersubunit interactions upon substrate engagement or recruitment of binding partners. A more detailed examination of the reaction kinetics would be required to determine, for example, whether all subunits in the hexameric ring of p97 are equally good substrates or whether subunits that lie along the 3-fold axis preferentially cross-link to yield dimers. Although in the crystal structure (28) all of the individual subunits appear identical, it remains to be determined whether this equivalency applies in solution as well. For cyclic proteins, there is compelling evidence that demonstrates improved stability of circularized proteins when compared with their linear counterparts (35–37, 40, 49). This is true for cyclic versions GFP (35),  $\beta$ -lactamase (36), and dihydrofolate reductase (37) generated using intein-based methods. The extension of protein cyclization to proteins of therapeutic value to improve the *in vivo* half-life has already been suggested (16, 39) and remains an exciting avenue for further research. Covalent closure of a protein through sortase-mediated circularization may also facilitate structural analysis of proteins whose flexible termini may interfere with crystallization.

*Acknowledgments*—We thank Carla Guimaraes and Mathias Pawlak for help in the preparation of the G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> and G-Cre-LPETG-His<sub>6</sub> substrates, respectively.

#### REFERENCES

- Marraffini, L. A., Dedent, A. C., and Schneewind, O. (2006) *Microbiol. Mol. Biol. Rev.* **70**, 192–221
- Paterson, G. K., and Mitchell, T. J. (2004) *Trends Microbiol.* **12**, 89–95
- Budzik, J. M., Marraffini, L. A., Souda, P., Whitelegge, J. P., Faull, K. F., and Schneewind, O. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10215–10220
- Budzik, J. M., Oh, S. Y., and Schneewind, O. (2008) *J. Biol. Chem.* **283**, 36676–36686
- Fälker, S., Nelson, A. L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Méléfors, O., Normark, S., and Henriques-Normark, B. (2008) *Mol. Microbiol.* **70**, 595–607
- Kruger, R. G., Otvos, B., Frankel, B. A., Bentley, M., Dostal, P., and McCafferty, D. G. (2004) *Biochemistry* **43**, 1541–1551
- Fallen, M. J., Lam, A. C., Antonio, M., and Dunbar, K. (2001) *Trends Microbiol.* **9**, 97–102
- Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12424–12429
- Schneewind, O., Fowler, A., and Faull, K. F. (1995) *Science* **268**, 103–106
- Ton-That, H., and Schneewind, O. (1999) *J. Biol. Chem.* **274**, 24316–24320
- Antos, J. M., Miller, G. M., Grotenbreg, G. M., and Ploegh, H. L. (2008) *J. Am. Chem. Soc.* **130**, 16338–16343
- Samantary, S., Marathe, U., Dasgupta, S., Nandicoori, V. K., and Roy, R. P. (2008) *J. Am. Chem. Soc.* **130**, 2132–2133
- Pritz, S., Wolf, Y., Kraetke, O., Klose, J., Bienert, M., and Beyertmann, M. (2007) *J. Org. Chem.* **72**, 3909–3912
- Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E., and Ploegh, H. L. (2007) *Nat. Chem. Biol.* **3**, 707–708
- Tanaka, T., Yamamoto, T., Tsukiji, S., and Nagamune, T. (2008) *ChemBiochem* **9**, 802–807
- Parthasarathy, R., Subramanian, S., and Boder, E. T. (2007) *Bioconjugate Chem.* **18**, 469–476
- Chan, L., Cross, H. F., She, J. K., Cavalli, G., Martins, H. F., and Neylon, C. (2007) *PLoS ONE* **2**, e1164
- Clow, F., Fraser, J. D., and Proft, T. (2008) *Biotechnol. Lett.* **30**, 1603–1607
- Mao, H., Hart, S. A., Schink, A., and Pollak, B. A. (2004) *J. Am. Chem. Soc.* **126**, 2670–2671
- Mao, H. (2004) *Protein Expr. Purif.* **37**, 253–263

## Chapter 4: A straight path to circular proteins

### Circular Proteins

21. Gillon, A. D., Saska, I., Jennings, C. V., Guarino, R. F., Craik, D. J., and Anderson, M. A. (2008) *Plant J.* **53**, 505–515
22. Saska, I., and Craik, D. J. (2008) *Trends Biochem. Sci.* **33**, 363–368
23. Saska, I., Gillon, A. D., Hatsugai, N., Dietzgen, R. G., Hara-Nishimura, I., Anderson, M. A., and Craik, D. J. (2007) *J. Biol. Chem.* **282**, 29721–29728
24. Peitz, M., Jäger, R., Patsch, C., Jäger, A., Egert, A., Schorle, H., and Edenhofer, F. (2007) *Genesis* **45**, 508–517
25. Popp, M. W., Artavanis-Tsakonas, K., and Ploegh, H. L. (2009) *J. Biol. Chem.* **284**, 3593–3602
26. Tsien, R. Y. (1998) *Annu. Rev. Biochem.* **67**, 509–544
27. Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Ploegh, H. L., and Kessler, B. M. (2002) *Chem. Biol.* **9**, 1149–1159
28. Davies, J. M., Brunger, A. T., and Weis, W. I. (2008) *Structure* **16**, 715–726
29. Martin, S. S., Pulido, E., Chu, V. C., Lechner, T. S., and Baldwin, E. P. (2002) *J. Mol. Biol.* **319**, 107–127
30. Misaghi, S., Galarzy, P. J., Meester, W. J., Ovaa, H., Ploegh, H. L., and Gaudet, R. (2005) *J. Biol. Chem.* **280**, 1512–1520
31. Yang, F., Moss, L. G., and Phillips, G. N., Jr. (1996) *Nat. Biotechnol.* **14**, 1246–1251
32. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132
33. Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003) *Biochemistry* **42**, 11307–11315
34. Ton-That, H., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (2000) *J. Biol. Chem.* **275**, 9876–9881
35. Iwai, H., Lingel, A., and Pluckthun, A. (2001) *J. Biol. Chem.* **276**, 16548–16554
36. Iwai, H., and Pluckthun, A. (1999) *FEBS Lett.* **459**, 166–172
37. Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C., and Benkovic, S. J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13638–13643
38. Love, K. R., Catic, A., Schlieker, C., and Ploegh, H. L. (2007) *Nat. Chem. Biol.* **3**, 697–705
39. Craik, D. J. (2006) *Science* **311**, 1563–1564
40. Trabi, M., and Craik, D. J. (2002) *Trends Biochem. Sci.* **27**, 132–138
41. Camarero, J. A., and Muir, T. W. (1997) *Chem. Commun.* 1369–1370
42. Camarero, J. A., Pavel, J., and Muir, T. W. (1998) *Angew. Chem.-Int. Ed. Engl.* **37**, 347–349
43. Daly, N. L., Love, S., Alewood, P. F., and Craik, D. J. (1999) *Biochemistry* **38**, 10606–10614
44. Jackson, D. Y., Burnier, J. P., and Wells, J. A. (1995) *J. Am. Chem. Soc.* **117**, 819–820
45. Hartgerink, J. D., Granja, J. R., Milligan, R. A., and Ghadiri, M. R. (1996) *J. Am. Chem. Soc.* **118**, 43–50
46. Camarero, J. A., Fushman, D., Sato, S., Giriat, I., Cowburn, D., Raleigh, D. P., and Muir, T. W. (2001) *J. Mol. Biol.* **308**, 1045–1062
47. Camarero, J. A., Kimura, R. H., Woo, Y. H., Shekhtman, A., and Cantor, I. (2007) *Chembiochem* **8**, 1363–1366
48. Evans, T. C., Jr., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, L., Chen, L., Benner, J., Liu, X. Q., and Xu, M. Q. (2000) *J. Biol. Chem.* **275**, 9091–9094
49. Colgrave, M. L., and Craik, D. J. (2004) *Biochemistry* **43**, 5965–5975

## Chapter 4: A straight path to circular proteins

### Supplemental Information

**Synthesis of Triglycine Tetramethylrhodamine Peptide.** GGG-TMR was synthesized using standard Fmoc solid-phase peptide synthesis (supplemental Fig. S7). Resin bound intermediate S1 was synthesized on Rink amide resin using reported methods (1). Resin S1 (72 mg, 0.036 mmol, 0.5 mmol/g resin loading) was then treated with 2 mL of 94:5:1 CH<sub>2</sub>Cl<sub>2</sub>/TIPS/TFA (5x, ~5 min each) to remove the 4-methyltrityl (Mtt) protecting group. The resin was then washed with 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x each, 3-5 min each wash) followed by 2.5 mL of NMP (3x each, 3-5 min each wash). The resin was then treated with a solution of 5(6)-TAMRA-NHS (25 mg, 0.047 mmol, mixture of 5 and 6 isomers, C1171 Invitrogen) and DIPEA (25 μL, 0.14 mmol) in 1 mL of NMP. The reaction mixture was incubated at RT for 19 h with gentle agitation. The resin was then washed with 2.5 mL of NMP (3x each, 3-5 min each wash) followed by 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x each, 3-5 min each wash). The resin was then deprotected with 2.5 mL of 80:20 NMP/piperidine for 20 min at RT followed by washing with 2.5 mL of NMP (3x each, 3-5 min each wash) and 2.5 mL of CH<sub>2</sub>Cl<sub>2</sub> (3x each, 3-5 min each wash). The peptide was then cleaved from the resin with 2.5 mL of 95:3:2 TFA/TIPS/H<sub>2</sub>O (5x, 15-20 min each). The combined cleavage solutions were concentrated and the crude peptide was precipitated from cold diethyl ether. The individual isomers of GGG-TMR were finally purified by RP-HPLC on a Waters Delta Pak 15 μm, 100 Å C18 column (7.8 x 300 mm, MeCN:H<sub>2</sub>O gradient mobile phase containing 0.1% TFA, 3 mL/min). Two major peaks of roughly equal abundance were collected and gave the expected mass for GGG-TMR when analyzed by ESI-MS (LRMS, calculated for GGG-TMR [M<sup>+</sup>] = 729.3, found 729.4 for both isomers). The individual fractions were not assigned as either the 5 or 6 isomers. Both isomers of GGG-TMR were used interchangeably in sortase-catalyzed transpeptidation reactions with no effect on protein labeling efficiency.

### References

1. Antos, J. M., Miller, G. M., Grotenbreg, G. M., and Ploegh, H. L. (2008) *J. Am. Chem. Soc.* **130**, 16338-16343

## Chapter 4: A straight path to circular proteins

### Supplemental Figure Legends

**Supplemental Figure S1.** Characterization of cyclic eGFP. (a) ESI-MS of linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> and circular eGFP formed by intramolecular transpeptidation. (b) A variant of eGFP lacking N-terminal glycine residues yields a mixture of unreacted starting material and hydrolysis when treated with sortase A. Conditions: eGFP-LPETG-His<sub>6</sub> (36 μM) and sortase A (50 μM) in (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 23 hours at 37 °C. This preparation of eGFP-LPETG-His<sub>6</sub> consisted of a mixture in which a fraction of the protein had the N-terminal Met residue cleaved during the course of protein expression to reveal an N-terminal Val residue. (c) MS/MS spectrum of a tryptic fragment of circular eGFP showing ligation of the N-terminal glycine residues (shown in red) to the C-terminal LPET residues (shown in green). Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. The full amino acid sequence of linear G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub> is shown at the top.

**Supplemental Figure S2.** Characterization of cyclic UCHL3 by MS/MS analysis unambiguously confirms the structure of the intramolecular transpeptidation product. The circular UCHL3 fragment was resolved by SDS-PAGE and the band corresponding to the putative circularization product was excised and digested with Glu-C. The resulting peptides were analyzed by LC-MS/MS. A peptide representing the junction between the UCHL3 N-terminus and the LPETG motif was identified by LC-MS/MS analysis. Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. The full amino acid sequence of linear G-His<sub>6</sub>-UCHL3-LPETG-XX is shown at the top.

**Supplemental Figure S3.** Intermolecular versus intramolecular transpeptidation for UCHL3. (a) UCHL3 cyclization is reversible. UCHL3 (30 μM) was circularized by the addition of sortase A (150 μM) for 3 hours. Subsequently, GGG (90 mM) was added to each reaction and incubated for the indicated times. Samples were analyzed by SDS-PAGE (top) or ESI-MS (bottom). (b) A large molar excess of GGG nucleophile is required to compete with the intramolecular UCHL3 circularization. UCHL3 (30 μM) and sortase A (150 μM) were incubated with the indicated concentrations of GGG nucleophile in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 3 hours at 37 °C, and analyzed by SDS-PAGE followed by coomassie staining.

**Supplemental Figure S4.** Circular UCHL3 is competent to react with a ubiquitin suicide probe. UCHL3 (30 μM) was first circularized (lane 5) by treatment with sortase A (150 μM) or incubated with 90 mM GGG and sortase A (150 μM) to yield the linear transpeptidation product (lane 7) for 3 hours at 37 °C. Samples were then withdrawn, diluted two-fold in labeling buffer (100 mM Tris, pH 8.0, 150 mM NaCl) and incubated with 4 μg of HA-UbVME for an additional hour at 25 °C. Reactions were quenched with sample buffer, analyzed by 12.5% SDS-PAGE, and either stained with coomassie (top) or transferred to nitrocellulose for anti-HA immunoblot (bottom).

**Supplemental Figure S5.** C-terminal labeling of p97. (a) G-His<sub>6</sub>-p97-LPSTG-XX (1.5 mg/mL) and G-His<sub>6</sub>-p97 (1.5 mg/mL) were incubated with sortase A (50 μM) and GGG-TMR (1 mM) in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 2.5 h at 37 °C. SDS-PAGE revealed the expected C-terminal transpeptidation product for G-His<sub>6</sub>-p97-LPSTG-XX only.

## Chapter 4: A straight path to circular proteins

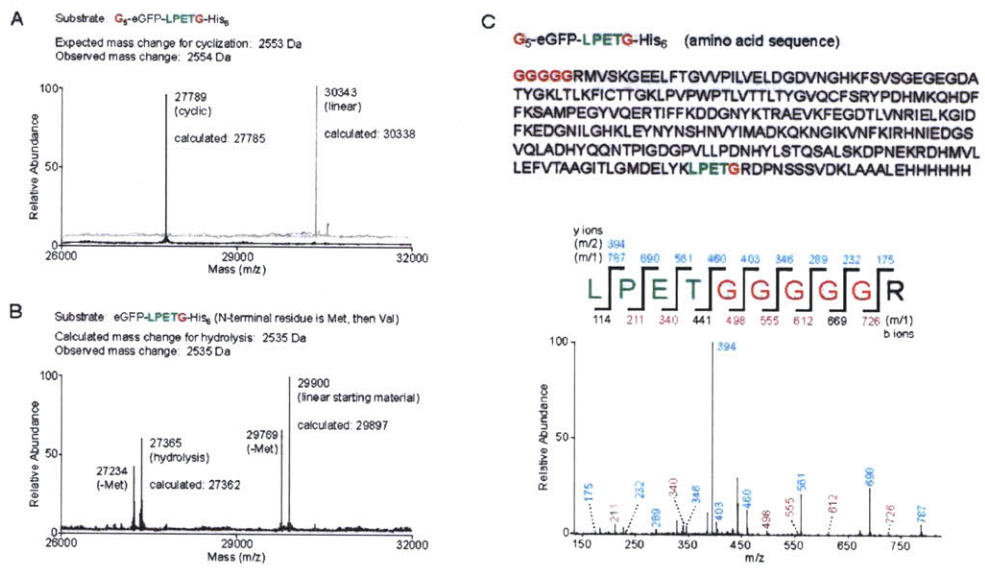
**Supplemental Figure S6.** G-His<sub>6</sub>-p97-LPSTG-XX was incubated with sortase A in sortase reaction buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>) for 2 hours at 37 °C. The extent of oligomerization at this early time point was much less than that seen after overnight incubation. The reaction mixture was then separated by SDS-PAGE. The band corresponding to dimeric p97 was excised, digested with chymotrypsin, and the resulting peptides analyzed by MS/MS. A peptide fragment showing the ligation of the N-terminal glycine residue (shown in red) to the C-terminal LPST residues (shown in green) was identified. Expected masses for y and b ions are listed above and below the peptide sequence. Ions that were positively identified in the MS/MS spectrum are highlighted in blue or red. Only the most prominent daughter ions have been labeled in the MS/MS spectrum. The full amino acid sequence of linear G-His<sub>6</sub>-p97-LPSTG-XX is shown at the top.

**Supplemental Figure S7.** Synthesis of GGG-TMR.

**Supplemental Figure S8.** Protein sequences.

*Chapter 4: A straight path to circular proteins*

Supplemental Figure S1



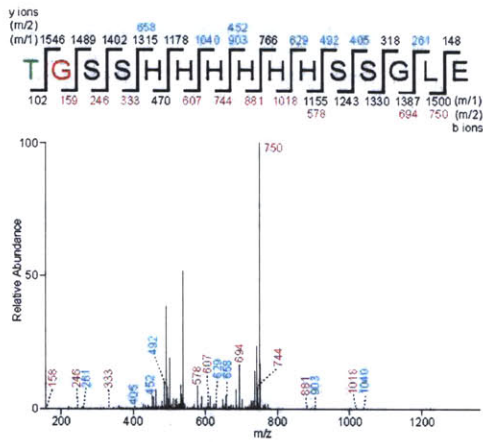
Downloaded from www.jbc.org at Inel Trinity College Dublin, on March 30, 2011

*Chapter 4: A straight path to circular proteins*

Supplemental Figure S2

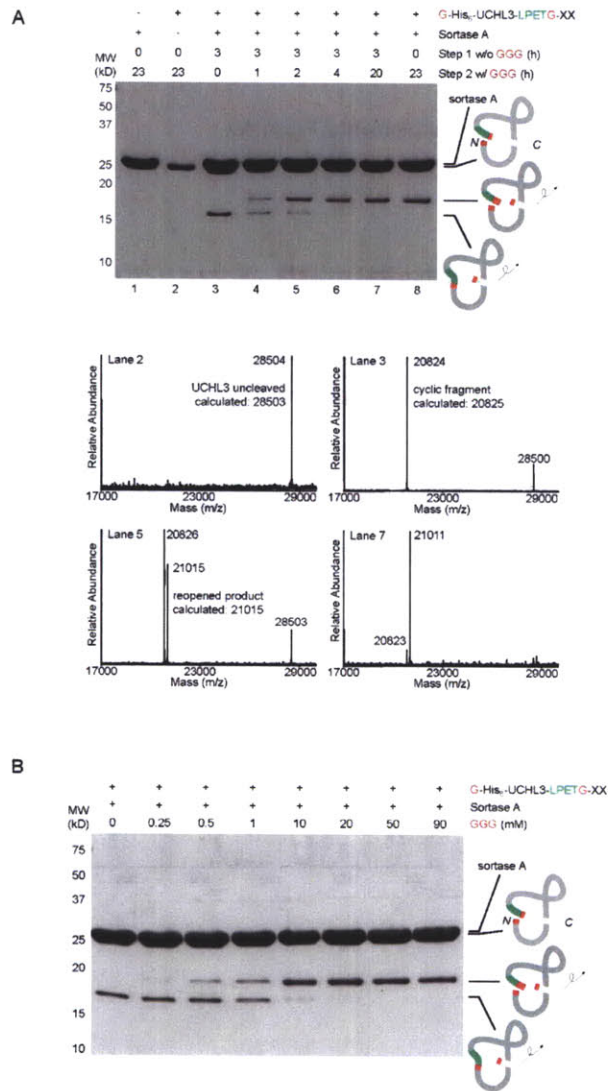
**G-His<sub>6</sub>-UCHL3-LPETG-XX** (amino acid sequence)

GSSHHHHHSSGLEVLFGPHMEGQRWLPLEANPEVTNQFLKQ  
 LGLHPNWQFVDVYGMPELLSMVPRPVCALLLPITEKYEVFRT  
 EEEEKIKSQGQDVTSSVYFMKQTSINACGTIGLIHAIANNKDKMHF  
 ESGSTLKKFLEESVSMSPPEARARYLENYDAIRVTHETSAHEGQTE  
 LPETG**E**KVDLHFIALVHVDGHLVELDGRKPPINHGSETSDETLLED  
 AIEVCKKFMERDPDELRFNAIALSAA



*Chapter 4: A straight path to circular proteins*

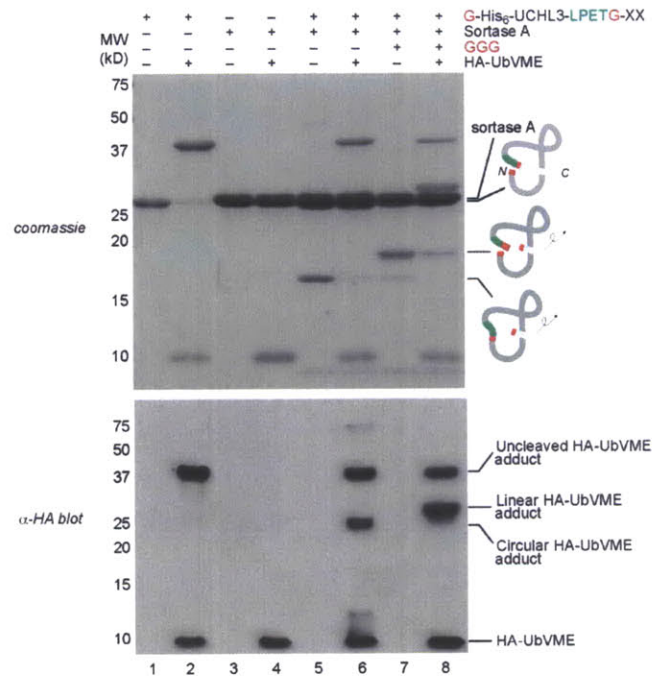
Supplemental Figure S3





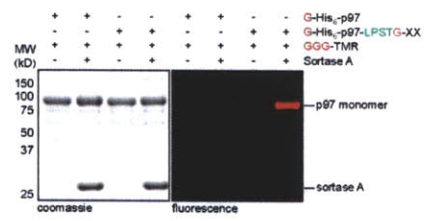
*Chapter 4: A straight path to circular proteins*

Supplemental Figure S4



Downloaded from www.jbc.org at Frel, Trinity College Dublin, on March 30, 2011

Supplemental Figure S5

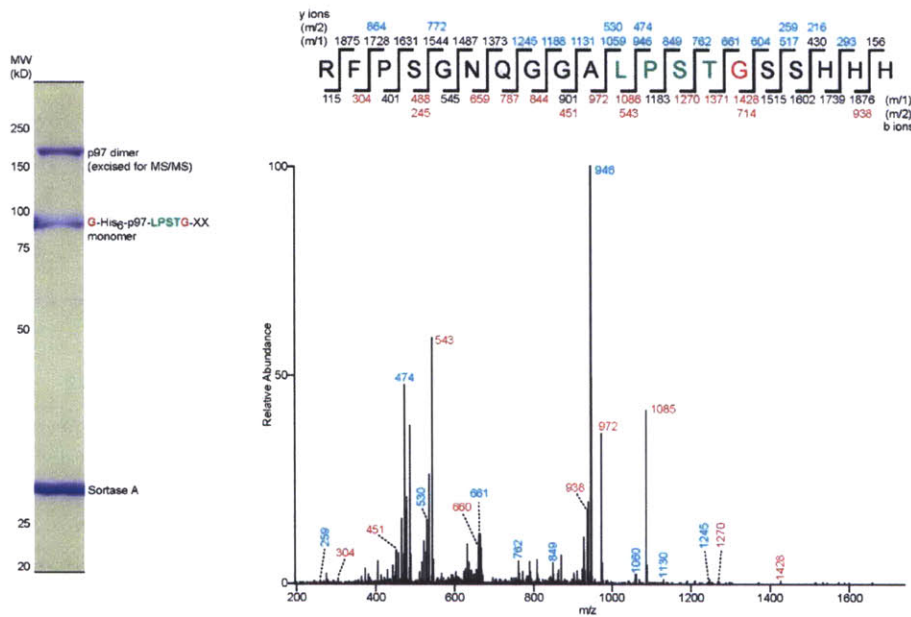


Chapter 4: A straight path to circular proteins

Supplemental Figure S6

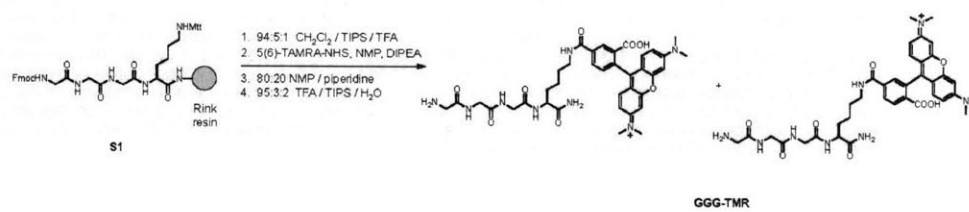
G-His<sub>6</sub>-p97-LPSTG-XX (amino acid sequence)

GSSHHHHHSSGLELVFQGPHMASGADSKGDDLSTAILKQKNRPNRLIVDEAINEDNSVVSLSQPKMDELQFRGDTVLLKGGKREAVCIVLS  
 DDTCSDKIRMNRVVRNLRVRLGDVSIQPCPDVKYGKRIHVLPIDDTVEGITGNLFEVYLKPYFLEAYRPIRKGDFLVRGGMRAVEFKVVEDP  
 SPYCVAPDTVHCEGEPIKREDEEESLNEVGYDDIGGCRKQLAQIKEMVELPLRHPALFKAIGVKPPRGILLYGPPGTGKTLIARAVANETGAFFL  
 INGPEIMSKLAGESNLKAFEEAEKNAPAIIFIDELDAIPKREKTHGEVERRIVSLLTMDGLKQRAHVIVMAATNRPNISIDPALRRFGRFDRE  
 VDIGIPDATGRLEILQIHTKNMKLADDVDLEQVANETHGHVGDALAEALQAIKMDLIDLEDETTDAEVMNSLAVTMDDFRWALSQSNPS  
 ALRETVVEVPQVTWEDIGGLEEDVKRELQELVQYPVEHPDKFLKFGMTPSKGVLFYGGPGCKTLLAKAIANECQANFISIKGPELLTMWVGESEA  
 NVREIFDKARQAAPCVLFFDELDSIAKARGGNIIDGGGAADRINQILTEMDGMSTKKNVFIIGATNRPDIIDPAILRPGRLDQLIYIPLPEKSRVAIL  
 KANLRKSPVAKVDVLEFLAKMTNGFSGADLTEICQRACKLAIESIESEIRRRERQTNPSAMEVEEDDPVPEIRRDHFEEAMRFARRSVSDNDIR  
 KYEMFAQLTQQSRGFGSFRFPSTGSGGG



Downloaded from www.jbc.org at IReL (Trinity College Dublin), on March 30, 2011

Supplemental Figure S7



*Chapter 4: A straight path to circular proteins*

**Supplemental Figure S8**

Sortase A:

MRGSHHHHHHSGSKPHIDNYLHDKDKDEKIEQYDKNVKEQASKDKKQQAQKPIPKDKS  
KVAGYIEIPDADIKVPYYPGPATPEQLNRGVSAEENASLDDQNISIAGHTFIDRPNYQFT  
NLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQKQKDKQLTLITCDDYNE  
KTGVWEKRKIFVATEVK

G-Cre-LPETG-His<sub>6</sub>:

GEFAPKKRKVSNLLTVHQNLPALPVDATSEVRKNLMDMFRDRQAFSEHTWKMLLS  
VCRSAAWCKLNNRWFPAPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGL  
PRPSDSNAVSLVRRIRKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLA  
FLGIAYNTLLRIAIAIRIRVKDISRTDGGRRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERW  
ISVSGVADDPNNYLFRCVRKNGVAAPSAATSQLSTRALEGIFEATHRLIYGAKDDSGQRY  
LAWSGHSARVGAARDMARAGVSIPEIMQAGGWTVNIVMNYIRNLDSETGAMVRLLO  
DGDTSGGGGSGGGGSASLPETGLEHHHHHHHH

G<sub>5</sub>-eGFP-LPETG-His<sub>6</sub>:

GGGGRRMVSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDATYGLTLKFICTTG  
KLPVPWPTLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR  
AEVKFEGDTLVNRIELKIGDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRH  
NIEDGSQLADHYQQNTPIGDGPVLLPDNHLYLSTQSALSKDPNEKRDMVLEFVTA  
GITLGMDELYKLPETGRDPNSSVDKLAALAEHHHHHH

eGFP-LPETG-His<sub>6</sub>

MVSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDATYGLTLKFICTTGKLPVPWP  
TLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEG  
DTLVNRIELKIGDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDG  
SQLADHYQQNTPIGDGPVLLPDNHLYLSTQSALSKDPNEKRDMVLEFVTAAGITLGM  
DELYKLPETGRDPNSSVDKLAALAEHHHHHH

G-His<sub>6</sub>-UCL3-LPETG-XX:

GSSHHHHHSSGLEVLFGQPHMEGQRWLPLEANPEVTNQFLKQLGLHPNWQFVDVY  
GMDPELLSMVPRPVCVAVLLLFPITEKYEVRTEEEEKIKSQGQDVTSSVYFMKQTISNA  
CGTIGLIHAIANNKDKMHFESGSLKFLFEEVSMSPPEARARYLENYDAIRVTHETSAHE  
GQTELPETGEKVDLHFIALVHVDGHLVELDGRKPPFPINHGETSDETLLEDAIEVCKKFME  
RDPDELRFNAIALSAA

*Chapter 4: A straight path to circular proteins*

G-His<sub>6</sub>-p97:

GSSHHHHHSSGLEVLFGQPHMASGADSKGDDLSTAILKQKNRPNRLIVDEAINEDNS  
VVSLSQPKMDELQLFRGDTVLLKGKKRREAVCIVLSDDTCSDEKIRMNRVVRNLRVR  
LGDVISIQPCPDVKYGKRIHVLPIDDTVEGITGNLFEVYLKPYFLEAYRPIRKGDIPLVRGG  
MRAVEFKVVEDPSPYCVAPDTVIHCEGEPIKREDEEESLNEVGYYDDIGGCRKQLAQIK  
EMVELPLRHPALFKAIGVKPPRGILLYGPPGTGKTLIARAVANETGAFFFLINGPEIMSKL  
AGESESNLRKAFEEAEKNAPAIIFIDELDAIAPKREKTHGEVERRIVSQQLLTMDGLKQRA  
HVIVMAATNRPNNSIDPALRRFRGFRDREVDIGIPDATGRLEILQIHTKNMKLADDVDLEQV  
ANETHGHVGDALAALCSEAALQAIRKKMDLIDLEDETIDAEMNSLAVTMDDFRWALSQ  
SNPSALRETVVEVPQVTWEDIGGLEVDKRELQELVQYPVEHPDKFLKFGMTPSKGVLF  
YGGPGCGKTLAKAIANECQANFISIKGPELLTMWFGSEANVREIFDKARQAAPCVLFF  
DELDSIAKARGGNIGDGGGAADRVINQILTEMDGMSTKKNVFIIGATNRPDIIDPAILRPG  
RLDQLIYIPLPDEKSRVAILKANLRKSPVAKDQVDLEFLAKMTNGFSGADLTEICQRACKLA  
IRESIESEIRREERERTNPSAMEVEEDDPVPEIRRDHFEEAMRFARRSVSDNDIRKYEM  
FAQTLQQSRGFGSFRFPNQGAGPSQGGGGTGGSVYTEDNDDDLYG

G-His<sub>6</sub>-p97-LPSTG-XX:

GSSHHHHHSSGLEVLFGQPHMASGADSKGDDLSTAILKQKNRPNRLIVDEAINEDNS  
VVSLSQPKMDELQLFRGDTVLLKGKKRREAVCIVLSDDTCSDEKIRMNRVVRNLRVR  
LGDVISIQPCPDVKYGKRIHVLPIDDTVEGITGNLFEVYLKPYFLEAYRPIRKGDIPLVRGG  
MRAVEFKVVEDPSPYCVAPDTVIHCEGEPIKREDEEESLNEVGYYDDIGGCRKQLAQIK  
EMVELPLRHPALFKAIGVKPPRGILLYGPPGTGKTLIARAVANETGAFFFLINGPEIMSKL  
AGESESNLRKAFEEAEKNAPAIIFIDELDAIAPKREKTHGEVERRIVSQQLLTMDGLKQRA  
HVIVMAATNRPNNSIDPALRRFRGFRDREVDIGIPDATGRLEILQIHTKNMKLADDVDLEQV  
ANETHGHVGDALAALCSEAALQAIRKKMDLIDLEDETIDAEMNSLAVTMDDFRWALSQ  
SNPSALRETVVEVPQVTWEDIGGLEVDKRELQELVQYPVEHPDKFLKFGMTPSKGVLF  
YGGPGCGKTLAKAIANECQANFISIKGPELLTMWFGSEANVREIFDKARQAAPCVLFF  
DELDSIAKARGGNIGDGGGAADRVINQILTEMDGMSTKKNVFIIGATNRPDIIDPAILRPG  
RLDQLIYIPLPDEKSRVAILKANLRKSPVAKDQVDLEFLAKMTNGFSGADLTEICQRACKLA  
IRESIESEIRREERERTNPSAMEVEEDDPVPEIRRDHFEEAMRFARRSVSDNDIRKYEM  
FAQTLQQSRGFGSFRFPNQGALPSTGSGGG

## **Part II. Applications of the sortase methodology**

**Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations**



**Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations**

(from: M. W. Popp, K. Artavanis-Tsakonas, H. L. Ploegh, *J Biol Chem* **2009**, 284, 3593)

**Abstract**

Determining how deubiquitinating enzymes discriminate between ubiquitin-conjugated substrates is critical to understand their function. Through application of a novel protein cleavage and tagging technique, sortagging, we show that human UCHL3 and the *Plasmodium falciparum* homologue, members of the ubiquitin c-terminal hydrolase family, use a unique active site crossover loop to restrict access of bulky ubiquitin adducts to the active site. Although it provides connectivity for critical active site residues in UCHL3, physical integrity of the crossover loop is dispensable for catalysis. By enlarging the active site crossover loop, we have constructed gain-of-function mutants that can accept substrates that the parent enzyme cannot, including ubiquitin chains of various linkages.

**Introduction**

Covalent posttranslational modification of proteins with the 76 amino-acid ubiquitin (Ub)<sup>1</sup> molecule controls many cellular processes, including protein turnover, trafficking,

---

<sup>1</sup> The abbreviations used are: Ub, ubiquitin; Ubl, ubiquitin-like molecule; UCH, ubiquitin C-terminal hydrolase; DUB, deubiquitinating enzyme; JAMM, JAB1/MPN/Mov34 metalloenzyme; OTU, ovarian tumor domain; USP, ubiquitin specific protease; MJD, Machado-Joseph disease protein domain; Fmoc, N-(9Fluorenylmethoxycarbonyl); Ub-AMC, Ub c-terminal 7-amido-4-methylcoumarin; HA-UbVME, Hemagglutinin epitope tagged ubiquitin vinylmethyl ester; HRP, horseradish peroxidase; LC-ESI-MS, liquid chromatography- electrospray ionization-mass spectrometry; PAGE, polyacrylamide gel electrophoresis; PVDF, polyvinylidene fluoride; PBS, phosphate buffered saline; Ni-NTA, nickel-nitrilotriacetic acid; M48, murine cytomegalovirus M48; PfUCHL3, *Plasmodium falciparum* UCHL3; DTT, dithiothreitol; SrtA, sortase A; Iso-T, Isopeptidase-T (USP5); HECT, homologous to E6-Associated Protein (E6AP) C-terminus; Mcl-1, myeloid cell leukemia-1 protein; Bcl-2, B-cell leukemia/lymphoma 2

and transcriptional regulation<sup>1</sup>. Ubiquitin conjugation to the  $\alpha$ -amine of lysine residues in target proteins is controlled by a series of enzymes: ubiquitin activating enzyme (E1), ubiquitin-conjugating enzymes (E2), and ubiquitin ligases (E3). Ubiquitin may also be condensed to other ubiquitin molecules, predominantly via internal lysines at positions 48 (K48) or 63 (K63) to form ubiquitin chains. The ubiquitin protein is generated as a head-to-tail fusion from the Ubb and Ubc loci and as a linear fusion to the ribosomal protein, CEP52. Like other regulatory modifications, ubiquitination is reversible. Removal of ubiquitin is the purview of deubiquitinating enzymes (DUBs), comprised of five groups: JAMM motif proteases, ovarian tumor proteases (OTUs), Ubiquitin specific protease (USPs), Machado-Joseph disease protein domain proteases (MJDs), and Ubiquitin C-terminal hydrolases (UCHs)<sup>2-3</sup>. Substrate recognition by these proteases is not well understood and it is highly likely that domains outside of the minimal catalytic unit regulate it. Save for members of the UCH class, no other DUBs have been crystallized in their full-length form. The UCH enzymes are proteins of modest size, capable of hydrolyzing ubiquitin adducts with small leaving groups, and contribute to homeostasis of ubiquitin levels in the cell. It is widely held that many members of this class of ubiquitin-specific hydrolases are unlikely to be involved in editing of ubiquitin-modified proteins, but rather recycle ubiquitin that has been consumed by reactions with small molecules<sup>3</sup>. Accordingly, large N-terminal ubiquitin fusion proteins are generally poor substrates for UCH proteases *in vitro*<sup>4</sup>. Since detailed structural data is available for several members of the UCH class of DUBs in their full-length form, we chose to study

---

protein; *E. coli*, *Escherichia coli*; ISG15, Interferon stimulated gene 15; CEP52, 52 amino acid 60S ribosomal protein (L40)

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

substrate recognition by these enzymes.

The structures of the yeast UCHL3 homologue YUH-1 <sup>5</sup>, as well as those of the mammalian UCHL1 <sup>6</sup> and UCHL3 enzymes are known, for UCHL3 both in its free <sup>7</sup> and substrate-occupied form <sup>8</sup>. UCHL3 is an unusual enzyme from a topological perspective: it possesses a highly knotted structure, possibly an evolutionary solution to survival in the proteolytic environment of the ubiquitin-proteasome system <sup>9</sup>. A distinguishing feature of the enzyme's architecture is the presence of an active site crossover loop that embraces the C-terminal segment of the ubiquitin suicide substrate with which the enzyme was co-crystallized <sup>8</sup>. In the absence of substrate, the crossover loop is flexible and not visible in the X-ray structure. The role of this loop, and the relevance of its movements in the course of catalysis is unclear, but it has been proposed that this loop aids in the proper positioning of the substrate -a ubiquitin adduct- in the enzyme's active site. In contrast to UCHL3, analysis of the UCHL1 crystal structure (51% sequence identity to UCHL3) reveals occlusion of the active site by a crossover loop that is ordered also in the absence of ubiquitin <sup>6</sup>, but perhaps this is because of crystal packing interactions. The comparison of the UCHL1 and UCHL3 enzymes thus leaves the role of the crossover loop in catalysis or positioning of the substrate unresolved.

We engineered a cleavage site in the crossover loop of UCHL3, to explore its contribution to both structure and function. We chose to install a sortase recognition site, because it allows a site-specific cleavage and trans-acylation reaction with concomitant installation of a functionality (biotin, fluorophores) at the site of sortase cleavage <sup>10</sup>. By

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

applying the sortagging technique, we can simultaneously interrupt the connectivity of a protein's peptide backbone and install a tag to track only the cleaved species. Thus, both native and cleaved sortase substrates can be tracked simultaneously in the same reaction mixture. The properties of sortagged UCHL3 inspired us to introduce yet other alterations in the crossover loop, resulting in gain-of-function mutants of UCHL3.

## Results

*Engineering a sortase cleavage site into UCHL3*- The presence of the unusual active site crossover loop in UCHL3 suggests that it may play a role in substrate selection by making sequence-specific contacts to the substrate: either to the ubiquitin (Ub) or ubiquitin-like (Ubl) moiety or to the attached leaving group. Alternatively, the crossover loop may play a role in stabilization of the catalytic center of the enzyme. The catalytic cysteine (C95) and oxyanion-hole stabilizing glutamine (Q89) are separated from the general base histidine (H169) and aspartic acid (D184) residues by the crossover loop, with H169 lying only three residues from the C-terminal end of the loop. The crossover loop not only traverses the active site of the enzyme, but also provides connectivity for the two halves of the catalytic center (**Figure 5.1a, Figure 5.4a**). The loop thus bridges key residues, bringing them into proximity for catalysis and possibly imparting stability to the active site. Alternatively, the active site crossover loop has been suggested to act as a substrate filter, limiting the size of the ubiquitin C-terminal leaving group<sup>7</sup>, a possibility that has been suggested but never experimentally addressed. To examine these possibilities, we engineered an LPETG sortase cleavage site into the crossover loop of both human UCHL3 and the *Plasmodium falciparum* homolog, PfUHCL3. We

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

included PfUCHL3 in our analysis because the two enzymes are structurally similar<sup>2</sup>, yet possess completely unrelated sequences in their crossover loop. As will be described below, results obtained for UCHL3 and PfUCHL3 are largely similar and thus allow generalization of our conclusions.

We inserted the LPETG cleavage site at three separate positions in the human UCHL3 (UCHL3) crossover loop and at two positions in the *Plasmodium falciparum* UCHL3 (PfUCHL3) crossover loop (**Supplementary Table 5.1, Figure 5.1a**), and expressed the mutant enzymes in *E. coli*. When exposed to Sortase A (SrtA) and the biotinylated oligoglycine nucleophile **1**, we observed successful transacylation for all mutant enzymes, albeit at different efficiencies (**Figure 5.1b**). The loop in both human and plasmodium UCHL3 therefore adopts a flexible conformation in solution. The extent of cleavage of the UCHL3 Loop 2 variant by SrtA can be modulated from 10% to nearly 90% by incubation with increasing amounts of SrtA and increasing time (**Supplementary Figure 5.1**). Ubiquitin hydrolase activity was largely unchanged, as assessed by hydrolysis of the ubiquitin-7-amino-4-methylcoumarin (Ub-AMC) substrate (**Figure 5.1c**). In addition, the LPETG substituted human UCHL3 Loop 2 mutant formed a covalent adduct with a hemagglutinin (HA)-tagged ubiquitin vinylmethyl ester suicide substrate (HA-UbVME), designed to react with the active site cysteine in DUBs, with similar efficiency

---

<sup>2</sup> Artavanis-Tsakonas, K., Weihofen, W. A., Gaudet, R., and Ploegh, H. L. (*manuscript in preparation*)

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

as did wild-type UCHL3 (**Figure 5.1d**). We chose to use the human UCHL3 Loop 2 and PfUCHL3 Loop 2 mutants for further analysis because these mutants retained near wild-type Ub-AMC hydrolysis activity and were most efficiently cleaved by sortase. We conducted subsequent sortagging experiments at low sortase concentrations that yield incomplete conversion to the transpeptidation product in order to track the properties of both cleaved and uncleaved UCHL3 species under identical conditions.

*Crossover loop connectivity is dispensable for ubiquitin docking-* The crossover loop in liganded UCHL3 likely assumes an ordered conformation in solution <sup>8</sup>. When purified UCHL3-HA-UbVME adduct is exposed to SrtA, no transpeptidation is observed (**Figure 5.2a**), consistent with previous observations that the LPETG motif must be placed in a flexible, unstructured region <sup>10</sup>. In addition, titration of ubiquitin into the sortase cleavage reaction successfully inhibits transpeptidation (**Figure 5.2b**). Therefore, in solution, the liganded form of UCHL3 possesses a rigid crossover loop refractory to attack by sortase. The highly knotted structure of UCHL3 suggests that it is possible to nick the crossover loop without complete unfolding of the polypeptide. We find only a slight increase in Stokes' radius upon cleavage of the crossover loop (**Figure 5.3**), suggesting that the nicked preparation does not undergo gross alterations in folding state. Nonetheless, the slightly larger hydrodynamic radius of the transpeptidation product is likely due to a more relaxed conformation caused by opening of the flexible crossover loop and installation of the 771 Da biotinylated probe (**Figure 5.3b**). We further examined whether connectivity of the active site is essential for maintenance of UCHL3 structure by *first* cleaving the active site loop with small amounts of sortase and *then*

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

delivering an HA-UbVME suicide substrate. Incomplete cleavage by sortase allows the properties of both the cleaved and uncleaved UCHL3 species to be examined simultaneously in one and the same reaction mixture. Indeed, a fraction of the cleaved UCHL3 successfully reacts with HA-UbVME, as shown by the appearance of a species that is both biotinylated and anti HA-reactive (**Figure 5.4b**). Since the HA-UbVME adduct is refractory to cleavage by sortase, sortase-mediated transpeptidation must have preceded HA-UbVME addition. Because of the electrophilicity of the vinyl methylester group, we predicted that HA-UbVME adduct formation would require only the cysteine nucleophile and not the general base histidine residue. We find that mutation of His169 to alanine results in HA-UbVME reaction levels comparable to wild-type UCHL3 (**Figure 5.4b, bottom**), but eliminates Ub-AMC hydrolysis (**Figure 5.1c**). Reaction of the catalytic cysteine with HA-UbVME requires proper folding of the enzyme to allow ubiquitin recognition, as evidenced by highly specific alkylation of DUB enzymes in whole cell lysate experiments<sup>11</sup>. Moreover, DUB protein preparations typically contain a small fraction of unfolded protein that does not react with electrophilic ubiquitin derivatives<sup>12</sup> (**Figure 5.1d**), despite the presence of a cysteine residue. Thus HA-UbVME reactivity reflects competency to bind ubiquitin and to position the catalytic cysteine in proximity of the electrophilic trap, and not ubiquitin hydrolase activity *per se*. Taken together, we conclude that connectivity of the active site crossover loop is not essential for maintenance of UCHL3 structure, as assessed by the ability of the cleaved preparation to interact with ubiquitin.

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

*The crossover loop restricts substrate size-* We reasoned that if the role of the active site crossover loop is to restrict the size of the ubiquitin C-terminal leaving group, then ablation of the connectivity should result in relaxed leaving group specificity. Indeed, UCHL3 with a nicked active site crossover loop hydrolyzes K63-linked ubiquitin chains to monomeric ubiquitin, while the uncleaved preparation and sortase itself fail to do so (**Figure 5.5a**). Incubation of UCHL3 with increasing concentrations of SrtA results in greater oligo-ubiquitin hydrolysis, indicating that it is the cleaved UCHL3 species that mediates liberation of mono-ubiquitin. This is consistent with the HA-UbVME labeling experiments and indicates that a portion of the cleaved UCHL3 species not only retains proper structure, but also ubiquitin hydrolase activity. The modest ubiquitin hydrolysis activity of the cleaved UCHL3 likely indicates that a fraction of the cleaved material is inactive upon transpeptidation, perhaps because the structure is subtly destabilized. Nevertheless, the gain-of-function associated with the cleaved UCHL3 preparation indicates that at least some fraction is competent to attack polyubiquitin chains.

*Expansion of the active site crossover loop results in gain-of-function mutants-* Based on these results, we reasoned that insertion of additional amino acids in the active site crossover loop to extend it would stabilize the enzyme, while still allowing the observed relaxed substrate specificity. Accordingly, we inserted 5 and 10 glycines in the active site crossover loop on the background of the Loop 2 LPETG mutation. Hydrolysis of Ub-AMC was affected only mildly for both mutants (**Figure 5.1c**). We then tested the ability of these mutants to hydrolyze ubiquitin chains of different linkages in the complete absence of sortase (**Figure 5.5b**). Both of the UCHL3 variants with 5 or 10



Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

added glycines in the crossover loop readily disassemble K63- and K48- linked ubiquitin chains, whereas the uncleaved UCHL3 does not. None of the mutants are able to hydrolyze a linear hexahistidine-tagged head-to-tail diubiquitin fusion, whereas Isopeptidase-T (USP5) efficiently does so. The PfUCHL3 Loop 2 mutant was also subjected to loop expansion and incubated with the various ubiquitin polymers. The PfUCHL3 loop-expanded mutants hydrolyze K48-, and to a lesser extent K63- linked ubiquitin, but not linear ubiquitin polymers (**Figure 5.5c**). We assessed the rates of catalysis of our engineered mutant forms of UCHL3 and PfUCHL3 and compared these to USP8<sup>13</sup> and the A20 N-terminal domain<sup>14-15</sup> (**Figure 5.6**). Quantitation of K48-linked diubiquitin hydrolysis yields rates intermediate between those of USP8 and the A20 N-terminal domain (**Table 5.1**). Thus, by expanding the length of the crossover loop, we have successfully created gain-of-function versions of UCHL3 and PfUCHL3 that can hydrolyze bulky substrates which the parent molecules are unable to attack.

We tested whether the loop-expanded versions of UCHL3 can hydrolyze the proximal ubiquitin from a ubiquitinated substrate, Mcl-1 (myeloid cell leukemia-1 protein). The HECT (Homologous to E6-Associated protein (E6AP) C-terminus) domain E3-ligase, ARF-BP1/Mule, targets the antiapoptotic Bcl-2 (B-cell leukemia/lymphoma 2 protein) family member Mcl-1 for ubiquitination at several internal lysines *in vitro*<sup>16</sup>. When such ubiquitinated Mcl-1 preparations<sup>2</sup> are exposed to the UCHL3 mutant with the 10 glycine expansion, the mono-ubiquitinated form of Mcl-1 is hydrolyzed (**Figure 5.7a**). In contrast to wild-type UCHL3, substrates enjoy unfettered access to the active site of the M48<sup>USP</sup> isopeptidase. The M48<sup>USP</sup> domain, which has activity against K63- and K48-

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

linked ubiquitin chains similar to that of the loop-expanded UCHL3 enzymes<sup>17</sup>, also efficiently hydrolyzes the mono-ubiquitinated Mcl-1 substrate. To ensure that the observed multi-(mono)ubiquitinated forms of Mcl-1 indeed represent a single ubiquitin condensed onto several Mcl-1 lysines and not a ubiquitin chain added to a single lysine, we performed the Mcl-1 ubiquitination reaction with a ubiquitin variant incapable of polyubiquitin chain formation because it lacks lysine residues. Although the Mcl-1 ubiquitination pattern of the 13 available lysines in Mcl-1 is different with this ubiquitin variant, the UCHL3 mutant with 10 added glycines again hydrolyzes all proximal ubiquitin moieties, as assessed by a dramatic increase in the amount of Mcl-1 backbone released after digestion (**Figure 5.7b**).

### **Discussion**

We took advantage of the known crystal structures for liganded and unliganded UCHL3 and applied the sortagging technique to a mammalian and apicomplexan representative of UCHL3. We chose to compare PfUCHL3 with its mammalian homolog because their sequences are no more than 35.9% identical, with almost no similarity in the crossover loop<sup>18</sup>. Thus, the sequence-specific contributions of the crossover loop to catalysis can be parsed out of its structural contributions by comparison of both the host and pathogen enzymes. We have meanwhile solved the PfUCHL3 structure and found the overall fold to be virtually identical to its human counterpart<sup>3</sup>. Only the unliganded form of UCHL3 is a substrate for transpeptidation- UCHL3 was not cleaved when modified with HA-UbVME prior to exposure to sortase. Combined, our results suggest that the crossover loop of UCHL3 is indeed a flexible structure, despite the ordered loop visible in the

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

crystal structure of its relative, UCHL1. Since there is no obvious sequence similarity between the crossover loops of UCHL3 and PfUCHL3, the presence, but not the exact sequence of the crossover loop is essential to substrate selection. This conclusion is supported also by the lack of effect of the introduction of an LPETG tag at different positions in the crossover loop. The loop alone determines the size of the leaving group that can be liberated from ubiquitin by UCHL3. Integrity of the loop is not required for catalysis by UCHL3 - the enzyme can function with its backbone cleaved very close to the active site. The stability of the enzyme is likely enhanced by the complex and knotted topology of the polypeptide backbone. There are many examples of proteins that are cleaved in unstructured loops; such cleavage is often required to generate an active enzyme from its inactive precursor, to expose the fusogenic properties of viral fusion proteins<sup>19</sup>, or to render active the bacterial toxins of the AB type such as cholera toxin and *E. coli* heat labile enterotoxin<sup>20</sup>. The cleavage of unstructured loops does not, as a rule, perturb secondary or tertiary structure for those proteins where the two forms can be examined independently. However, the processed proteins but not their precursors are often poised to undergo major structural rearrangements upon engagement of the appropriate counterstructures.

We have obtained gain-of-function mutants of UCHL3 by enlarging the active site crossover loop. These mutants successfully disassemble both K48 and K63  $\epsilon$ -amine linked ubiquitin polymers, while wild-type UCHL3, which has a crossover loop diameter of ~12-15 Angstroms, cannot. When known structures for K48- and K63- linked ubiquitin dimers are modeled in the solved UCHL3-UbVME structure (**Supplementary**

**Figure 2**), the proximal ubiquitin sterically clashes with UCHL3. To explain the observed hydrolysis of these substrates, the hinge region that contains the scissile Gly-Lys bond must be sufficiently flexible in solution to allow displacement of the proximal ubiquitin by UCHL3, consistent with previous observations<sup>21</sup>. Because K63 lies near the amino terminus of ubiquitin,  $\alpha$ -linked dimers are hypothesized to have an open conformation similar to that of K63 linked diubiquitin<sup>22</sup>. However, none of the UCHL3 or PfUCHL3 mutants can hydrolyze an  $\alpha$ -amine linked dimer. This does not result from a difference between the  $\alpha$  and  $\epsilon$  linkage itself, since UCHL3 can hydrolyze a ubiquitin fusion to the ribosomal protein CEP52, in  $\alpha$ -linkage<sup>4</sup>. Instead, the scissile bond in the ubiquitin head-to tail fusion is likely inaccessible to the UCHL3 active site cysteine. When the head-to tail ubiquitin dimer is modeled by ISG15, which has two tandem  $\alpha$ -linked Ubl domains, and fitted into the UCHL3-UbVME structure, the region where the scissile bond would be lies far outside of the active site groove (**Supplementary Figure 2**). Although the structural details of ISG15 likely differ from those of a ubiquitin head-to-tail fusion, we favor a model where the lack of cleavage by UCHL3 is due to differences in the conformation of K48-, K63- and  $\alpha$ -amine linked ubiquitin dimers in solution.

The promiscuity of the loop-expanded UCHL3 variants provides a rationale for the properties of UCHL3. In the absence of other protein domains that target the enzyme to particular ubiquitinated proteins or ubiquitin chains of specific linkages<sup>23</sup>, UCHL3 limits access to its active site to only those substrates that can pass through the narrow bore of

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

the crossover loop. Thus it is unlikely that wild-type UCHL3 directly targets larger, folded proteins for deubiquitination <sup>24</sup>.

In summary, we have reported a novel method for exploring the contribution of flexible loops to protein structure and function. By applying the sortagging technique to the active site crossover loop in UCHL3 as well as mutagenesis, we can examine the contributions of the loop to both the structure and function of the enzyme. We have expanded the range of substrates that UCHL3 can hydrolyze and provide strong support for the notion that the unusual active-site crossover loop functions as a substrate filter, limiting the types of substrates that the enzyme can hydrolyze. Such modified versions of UCHL3 may be useful as general ubiquitin releasing enzymes for the study and identification of ubiquitin adducts.

## **Methods**

*Reagents*- Probe **1** was synthesized by standard N-(9-Fluorenylmethoxycarbonyl) (Fmoc) based solid phase peptide chemistry as described <sup>10</sup>. Triglycine was purchased from Sigma. Antibodies were purchased from the following vendors: anti-Hemagglutinin (HA) tag (antibody 3F10-HRP; Horseradish Peroxidase), Roche; streptavidin-HRP, Amersham Biosciences; anti-FLAG tag M2 (Sigma); anti-ubiquitin (rabbit, Sigma); anti-Mcl-1 (Sigma); goat anti-rabbit-HRP (Southern Biotech). FLAG peptide was purchased from Sigma. Ub C-terminal 7-amido-4-methylcoumarin (Ub-AMC), ubiquitin lacking all lysines, K63-linked diubiquitin, K48-linked diubiquitin, UBE1, USP8, and A20 N-terminal domain were purchased from Boston Biochem. Purified ubiquitin was

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

purchased from Sigma. Hemagglutinin epitope tagged ubiquitin vinylmethyl ester (HA-UbVME) was generated as described previously<sup>25</sup>.

*Mass Spectrometry-* Liquid chromatography- electrospray ionization-mass spectrometry (LC-ESI-MS) was performed on a Micromass LCT mass spectrometer (Micromass MS Technologies, USA) and a Paradigm MG4 HPLC system equipped with a HTC PAL autosampler (Michrom BioResources, USA) and a Waters Symmetry 5  $\mu$ M C8 column (2.1 x 50 mm, MeCN:H<sub>2</sub>O (0.1% formic acid) gradient mobile phase, 150  $\mu$ l/min).

*Cloning and protein expression-* Sortase A was expressed and purified as described<sup>26</sup>. Murine cytomegalovirus M48 deubiquitinating enzyme domain (M48<sup>USP</sup>) was cloned and purified as described previously<sup>17</sup>. The C-terminal residues of the E3 ligase, ARF-BP1 (amino acids 4012-4374), UCHL3 and PfUCHL3 were cloned into pET28a+ (Novagen) with an N-terminal hexahistidine tag. The N-terminal residues of Mcl-1 (myeloid cell leukemia-1 protein, amino acids 1-327) were cloned as an N-terminal FLAG peptide fusion into the vector pET16b (Novagen) carrying an N-terminal His<sub>10</sub> tag. The final Mcl-1 construct consists of an N-terminal His<sub>10</sub> tag followed by a FLAG tag and Mcl-1<sub>1-327</sub>. UCHL3 point mutations and LPETG substitutions were generated by site-directed mutagenesis using a QuickChange kit (Stratagene). Loop expanded versions of UCHL3 and PfUCHL3 were generated by inserting 5 or 10 glycine residues immediately after the LPETG sequence. Hexahistidine tagged linear diubiquitin was also cloned into pET28a+. Proteins were expressed in BL-21 *E. coli* or Rosetta (FLAG-Mcl-1) DE3(PLysS) cells (Novagen), resuspended in lysis buffer (50 mM Tris, 150 mM NaCl,

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

10 mM imidazole, 10% glycerol, pH 7.2) and lysed by French press. Proteins were purified from clarified lysates with nickel-nitrilotriacetic acid agarose (Ni-NTA; Qiagen) and eluted in lysis buffer supplemented with 500 mM imidazole. Eluted fractions were dialyzed extensively to remove imidazole. Protein concentrations were determined by the Bradford method (Biorad).

*Immunoblotting-* Proteins were separated by Tris/Glycine SDS-PAGE or Tris/Tricine PAGE (Ub chain digestion experiments) and transferred to nitrocellulose membranes or Polyvinylidene fluoride (PVDF, ubiquitin chain digestion experiments). Membranes were blocked with 5% nonfat dried milk in phosphate buffered saline supplemented with Tween 20 (PBS, 0.1% Tween 20, pH 7.4) overnight at 4 °C or for 1 h at room temperature. Membranes were washed with PBST and incubated with the indicated antibodies for 1 h. Streptavidin-HRP and anti-HA (3F10-HRP) blots were then washed and developed with Western Lighting Chemiluminescence Reagent Plus (Perkin Elmer). All other blots were washed, incubated with a goat anti-rabbit-HRP conjugate, and developed.

*Sortase cleavage and gel-filtration of sortagging reactions-* Sortagging of UCHL3 (10  $\mu$ M) was performed by incubating with the indicated concentrations of sortase A (Srt A) in Srt buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>), and 5 mM probe 1 in a 25  $\mu$ l volume. Reactions were incubated at 37 °C for 2 h and halted with sample buffer. For gel filtration, a 1 ml reaction (5  $\mu$ M sortase) was purified by size exclusion

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

chromatography on S75 Sephadex resin using 20 mM Tris, 50 mM NaCl, pH 8.0 buffer as eluent. Fractions were collected and analyzed by SDS-PAGE.

*Kinetic measurements-* Ub-AMC assays were performed in assay buffer (50 mM Tris/HCl, 150 mM NaCl, 2 mM EDTA, 2 mM DTT, 1 mg/ml BSA [pH 7.5]) at 25 °C. Enzyme concentrations were determined by Bradford assay (Biorad). UCHL3 mediated Ub-AMC hydrolysis was performed in a total volume of 30 µl with 10 pM UCHL3 mutants and 62.5 nM Ub-AMC in a 384-well NUNC black plate. PfUCHL3 assays were performed with 25 pM enzyme and 125 nM UbAMC. Data was collected with a Spectramax M2 plate reader (Molecular Devices) with a 368 nm/467 nm filter pair and a 455 nm cutoff. Velocities were determined by fitting the initial linear data points to a least-squares regression line.

*HA-UbVME labeling and adduct purification-* Sortase cleavage reactions (5 µM sortase) of the indicated UCHL3 mutants were performed for 2 h at 37 °C and 5 µl was diluted with 7 µl of labeling buffer (20 mM Tris pH 8.0, 150 mM NaCl). DTT (1 mM final concentration) and 1.5 µg of HA-UbVME was added and reactions were incubated at 37 °C for 1 h. Reactions were halted with sample buffer and loaded onto 12.5% SDS-PAGE for analysis. For HA-UbVME adduct purification, a large-scale reaction (1 ml) was prepared, omitting sortase and purified by ion-exchange chromatography as described <sup>8</sup>.



Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

*Ubiquitin polymer digestion- K63- linked diubiquitin cleavage by sortagged UCHL3*

Loop 2 enzyme was performed by incubating 7  $\mu\text{g}$  of UCHL3 Loop 2 with either 2.25  $\mu\text{g}$  or 22.5  $\mu\text{g}$  of SrtA and 5 mM Probe 1 in Srt buffer for 2 h at 37  $^{\circ}\text{C}$ . K63- linked diubiquitin (0.5  $\mu\text{g}$ ) was then added as well as 1 mM DTT and incubated for 1 h at 37  $^{\circ}\text{C}$ . Reactions were halted with Tris/Tricine sample buffer and subjected to 12.5 % Tris/Tricine PAGE followed by anti-ubiquitin immunoblot. Digestion of ubiquitin chains by UCHL3 loop extension mutants was performed in Srt buffer with 180 ng of each enzyme, 1 mM DTT and 1  $\mu\text{g}$  of the indicated ubiquitin chain in a total volume of 7  $\mu\text{l}$ . Reactions were halted with Tris/Tricine sample buffer and loaded onto a 12.5% Tris/Tricine gel for anti-ubiquitin immunoblot analysis. For kinetic measurements, 500 nM of enzyme was incubated with 10  $\mu\text{M}$  of either K63- or K48- linked diubiquitin in buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 2 mM EDTA, 2 mM DTT). At the indicated times, samples were withdrawn, quenched with Tris/Tricine loading buffer and subjected to 10% Tris/Tricine PAGE. Gels were stained with colloidal coomassie and bands for K48 diubiquitin hydrolysis were quantiated with ImageJ. Velocities were determined by fitting the initial linear data points to a least-squares regression line. Higher order K63- linked ubiquitin conjugates visible in colloidal blue stained gels precluded quantiation of K63 diubiquitin hydrolysis.

*Mcl-1 ubiquitination and digestion-* FLAG-Mcl-1 (1  $\mu\text{g}$ ) was incubated with 100 ng human UBE1 (E1), 1  $\mu\text{g}$  UbcH7 (E2), 10  $\mu\text{g}$  Arf/BP1 (E3) and 100  $\mu\text{g}$  Ub with an ATP

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

regenerating system<sup>16</sup> for 90 min at room temperature<sup>3</sup>. Reactions were quenched with NET buffer (50 mM Tris-HCl [pH 7.4], 0.5% NP-40, 150 mM NaCl, 5 mM EDTA) and incubated overnight at 4 °C with anti-FLAG M2 antibody. Immune complexes were recovered with Protein G beads (Sigma), washed extensively with NET buffer, and eluted by incubating with FLAG peptide (250 µg/ml in Srt Buffer) at 25 °C with shaking for 30 min. Eluates were collected, pooled, divided, and incubated with 1 mM DTT and UCHL3 or PfUCHL3 mutant enzymes (800 ng) for 2 h at 37 °C. Samples were analyzed by SDS-PAGE (10%), transferred to nitrocellulose, and anti-Mcl-1 immunoblot was performed.

---

<sup>3</sup> Love, K. R., Sastry, R. K., Spooner, E., and Ploegh, H.L. (*manuscript in preparation*)

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

**References**

1. Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* **22**, 159-180 (2006).
2. Love, K.R., Catic, A., Schlieker, C. & Ploegh, H.L. Mechanisms, biology and inhibitors of deubiquitinating enzymes. *Nat Chem Biol* **3**, 697-705 (2007).
3. Nijman, S.M. et al. A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**, 773-786 (2005).
4. Larsen, C.N., Krantz, B.A. & Wilkinson, K.D. Substrate specificity of deubiquitinating enzymes: ubiquitin C-terminal hydrolases. *Biochemistry* **37**, 3358-3368 (1998).
5. Johnston, S.C., Riddle, S.M., Cohen, R.E. & Hill, C.P. Structural basis for the specificity of ubiquitin C-terminal hydrolases. *EMBO J* **18**, 3877-3887 (1999).
6. Das, C. et al. Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1. *Proc Natl Acad Sci U S A* **103**, 4675-4680 (2006).
7. Johnston, S.C., Larsen, C.N., Cook, W.J., Wilkinson, K.D. & Hill, C.P. Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution. *EMBO J* **16**, 3787-3796 (1997).
8. Misaghi, S. et al. Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. *J Biol Chem* **280**, 1512-1520 (2005).
9. Virnau, P., Mirny, L.A. & Kardar, M. Intricate knots in proteins: Function and evolution. *PLoS Comput Biol* **2**, e122 (2006).
10. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**, 707-708 (2007).
11. Ovaa, H. et al. Activity-based ubiquitin-specific protease (USP) profiling of virus-infected and malignant human cells. *Proc Natl Acad Sci U S A* **101**, 2253-2258 (2004).
12. Hassiepen, U. et al. A sensitive fluorescence intensity assay for deubiquitinating proteases using ubiquitin-rhodamine110-glycine as substrate. *Anal Biochem* **371**, 201-207 (2007).
13. Mizuno, E. et al. Regulation of epidermal growth factor receptor down-regulation by UBPY-mediated deubiquitination at endosomes. *Mol Biol Cell* **16**, 5163-5174 (2005).
14. Lin, S.C. et al. Molecular basis for the unique deubiquitinating activity of the NF-kappaB inhibitor A20. *J Mol Biol* **376**, 526-540 (2008).
15. Wertz, I.E. et al. De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. *Nature* **430**, 694-699 (2004).
16. Chen, D. et al. ARF-BP1/Mule is a critical mediator of the ARF tumor suppressor. *Cell* **121**, 1071-1083 (2005).
17. Schlieker, C. et al. Structure of a herpesvirus-encoded cysteine protease reveals a unique class of deubiquitinating enzymes. *Mol Cell* **25**, 677-687 (2007).
18. Frickel, E.M. et al. Apicomplexan UCHL3 retains dual specificity for ubiquitin and Nedd8 throughout evolution. *Cell Microbiol* **9**, 1601-1610 (2007).

19. Skehel, J.J., Cross, K., Steinhauer, D. & Wiley, D.C. Influenza fusion peptides. *Biochem Soc Trans* **29**, 623-626 (2001).
20. Chinnapen, D.J., Chinnapen, H., Saslowsky, D. & Lencer, W.I. Rafting with cholera toxin: endocytosis and trafficking from plasma membrane to ER. *FEMS Microbiol Lett* **266**, 129-137 (2007).
21. Eddins, M.J., Varadan, R., Fushman, D., Pickart, C.M. & Wolberger, C. Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH. *J Mol Biol* **367**, 204-211 (2007).
22. Pickart, C.M. & Fushman, D. Polyubiquitin chains: polymeric protein signals. *Curr Opin Chem Biol* **8**, 610-616 (2004).
23. Reyes-Turcu, F.E., Shanks, J.R., Komander, D. & Wilkinson, K.D. Recognition of polyubiquitin isoforms by the multiple ubiquitin binding modules of isopeptidase T. *J Biol Chem* (2008).
24. Butterworth, M.B. et al. The deubiquitinating enzyme UCH-L3 regulates the apical membrane recycling of the epithelial sodium channel. *J Biol Chem* **282**, 37885-37893 (2007).
25. Borodovsky, A. et al. Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem Biol* **9**, 1149-1159 (2002).
26. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci U S A* **96**, 12424-12429 (1999).

## Figure Legends

### Figure 5.1. Characterization of UCHL3 and PfUCHL3 sortase substrates

(a) Crystal structure of UCHL3 in complex with UbVME (pdb: 1xd3). UbVME is in yellow with the crossover loop (blue) spanning the two halves of UCHL3 (green and orange). The position of the Loop 2 substitution is indicated by an arrow.

(b) LPETG substitution in the crossover loop renders UCHL3 and PfUCHL3 susceptible to sortase-mediated transpeptidation. Wild-type and LPETG substituted UCHL3 and PfUCHL3 mutants (10  $\mu$ M) were exposed to sortase (5  $\mu$ M top blot; 150  $\mu$ M bottom blot) and biotinylated oligoglycine probe 1 (5 mM) and analyzed by streptavidin-HRP immunoblot to detect the ~20 kD transpeptidation product.

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

(c) LEPTG-substituted UCHL3 and PfUCHL3 mutants catalyze Ub-AMC hydrolysis. Ub-AMC hydrolysis by wild-type and mutant UCHL3 (10 pM) was measured using sub-saturating concentrations of Ub-AMC (62.5 nM). Data is plotted as the ratio of velocity to the enzyme concentration, with  $n \geq 3$  (top). Enzymes with an H169A mutation are indicated by an (H), and enzymes with 5 or 10 glycines added to the crossover loop directly after the LPETG sequence are indicated by +5G and +10G. Ub-AMC hydrolysis (125 nM) was also measured for wild-type and mutant PfUCHL3 enzymes (25 pM). Data is plotted as the ratio of the velocity to the enzyme concentration, with  $n=3$  (bottom).

(d) An LPETG substituted UCHL3 mutant is labeled by a ubiquitin suicide substrate. HA-UbVME was titrated into a fixed amount (5  $\mu\text{g}$ ) of wild-type UCHL3 or the human UCHL3 Loop 2 mutant enzyme. Reactions were analyzed by 12.5% SDS-PAGE and stained with coomassie to visualize total protein.

**Figure 5.2. Ubiquitin binding alters susceptibility to sortase-mediated transpeptidation.**

(a) The UCHL3-HA-UbVME adduct is refractory to sortase-mediated transpeptidation. Free UCHL3 Loop 2 enzyme (10  $\mu\text{M}$ ) and the purified HA-UbVME adduct (10  $\mu\text{M}$ ) were exposed to SrtA (5  $\mu\text{M}$ ) and biotinylated probe 1 (5 mM) for the indicated times. Reactions were analyzed by 12.5% SDS-PAGE and either stained with coomassie (top) or transferred to nitrocellulose for streptavidin-HRP immunoblot (bottom).

(b) Ubiquitin binding competes with sortase-mediated transpeptidation. UCHL3 Loop 2 enzyme was incubated with SrtA (5  $\mu\text{M}$ ) and probe 1 (5 mM) for the indicated amounts of time in the absence or presence of various concentrations of purified ubiquitin. The transpeptidation product was detected by streptavidin-HRP immunoblot.

**Figure 5.3. Cleaved and intact UCHL3 migrate with similar Stokes' radius by gel filtration chromatography.**

(a) Similar gel filtration chromatography elution profiles of uncleaved UCHL3 Loop 2 alone (10  $\mu$ M); blue line and UCHL3 Loop 2 (10  $\mu$ M) incubated with SrtA (5  $\mu$ M) and probe 1 (5 mM); red line. Under these conditions, the majority of the UCHL3 population is uncleaved. Reactions were subjected to size exclusion chromatography on a Superdex S-75 column and the UV absorbance at 280 nm was recorded.

(b) Intact UCHL3 Loop 2 protein, cleaved UCHL3 Loop 2 protein, and Sortase A co-migrate by size exclusion chromatography. Fractions corresponding to the major peak (60 ml – 72 ml) in the sortase cleavage reaction (red line in **Fig. 3a**) where the majority of the UCHL3 population is uncleaved were collected and analyzed by 12.5 % SDS-PAGE followed by coomassie stain (top) to detect total protein or streptavidin-HRP immunoblot (bottom) to detect the transpeptidation product.

**Figure 5.4. Crossover loop integrity is not necessary for maintenance of UCHL3 active site structure.**

(a) Schematic showing the UCHL3 secondary structure elements with beta sheets represented as arrows and alpha helices as cylinders. The position of the crossover loop (blue) is shown relative to the catalytic residues (circled letters). Upon sortase-mediated transpeptidation, probe 1 is affixed to the half shown in green, interrupting the connectivity to the half shown in orange.

Note that the half of the enzyme that receives probe 1 contains the catalytic cysteine residue that reacts with HA-UbVME.

Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

(b) Cleaved UCHL3 preparations react with a ubiquitin suicide substrate. Wild-type UCHL3, UCHL3 Loop 2, and UCHL3 Loop2 H169A mutant enzymes (10  $\mu$ M) were exposed to SrtA (5  $\mu$ M) and probe 1 (5 mM) for 2 h at 37 °C and subsequently incubated with HA-UbVME (1.5  $\mu$ g) for 1 h at 37 °C. Reactions were analyzed by 12.5 % SDS-PAGE followed by streptavidin-HRP immunoblot and anti-HA immunoblot. The identities of the bands detected are shown (right).

**Figure 5.5. The active site crossover loop restricts ubiquitin leaving group size.**

(a) Cleavage of the crossover loop renders UCHL3 competent to disassemble ubiquitin polymers. UCHL3 Loop 2 protein (7  $\mu$ g) was exposed to either 2.25  $\mu$ g or 22.5  $\mu$ g of SrtA and probe 1 (5 mM) for 2 h. Reactions were subsequently incubated with K63- linked diubiquitin chains (0.5  $\mu$ g) at 37 °C for 1 h. Ubiquitin cleavage was assessed by 12.5% Tris/Tricine SDS-PAGE followed by anti-ubiquitin immunoblot. \*Ub<sub>n</sub> denotes higher order K63-linked ubiquitin in the diubiquitin preparation.

(b) Crossover loop expansion allows UCHL3-mediated hydrolysis of K63- and K48- linked ubiquitin but not a hexahistidine-tagged head-to-tail diubiquitin fusion. 180 ng of each enzyme (**2**= UCHL3 Loop 2, **5** = 5 glycine insertion, **10** = 10 glycine insertion, **Iso-T** = isopeptidase-T) was incubated with 1  $\mu$ g of the indicated ubiquitin polymer for 2 h at 37 °C. Reactions were separated by 12.5% Tris/Tricine SDS-PAGE followed by anti-ubiquitin immunoblot.. \*Ub<sub>n</sub> denotes higher order K63-linked ubiquitin in the diubiquitin preparation.

(c) Loop expanded versions of PfUCHL3 hydrolyze K48- linked diubiquitin, inefficiently cleave K63- linked polyubiquitin, and are inactive against a hexahistidine-tagged head-to-tail diubiquitin fusion. 180 ng of each enzyme (**2** = PfUCHL3 Loop 2, **5** = 5 glycine insertion, **10** =

10glycine insertion, **Iso-T** = isopeptidase-T) was incubated with 1  $\mu$ g of the indicated ubiquitin polymer for 2 h at 37 °C. Reactions were analyzed as in **Fig. 5b**

**Figure 5.6. Isopeptide linked diubiquitin hydrolysis by engineered UCHL3 and PfUCHL3 mutants.**

Enzymes (500 nM) were incubated with either K48- linked diubiquitin (10  $\mu$ M, left) or K63- linked diubiquitin (10  $\mu$ M, right), stopped at the indicated times points by addition of Tris/Tricine sample buffer, subjected to 10% Tris/Tricine SDS-PAGE electrophoresis, and stained by colloidal blue. K48- linked diubiquitin hydrolysis was quantified and rates were determined (**Table 1**). Because of visible higher order ubiquitin polymers, K63- linked ubiquitin hydrolysis is displayed, but not quantitated.

**Figure 5.7. Crossover loop expansion allows hydrolysis of ubiquitinated Mcl-1.**

(a) Loop expanded UCHL3 hydrolyzes ubiquitin-FLAG-Mcl-1 conjugates. FLAG-tagged Mcl-1 (1  $\mu$ g) was ubiquitinated by incubation with UBE1, UBCH7, ArfBP-1/Mule, and ubiquitin (see experimental procedures). Mcl-1 was then immunoprecipitated with anti-FLAG antibody and either boiled in sample buffer (input) or eluted with FLAG peptide (FLAG peptide eluate).

FLAG peptide eluates were pooled, divided equally, and incubated with 800 ng of the indicated enzymes for 2 h at 37 °C. Reactions were separated by 10% SDS-PAGE and analyzed by anti-Mcl-1 immunoblot.

(b) Loop expanded UCHL3 hydrolyzes the proximal ubiquitin from Mcl-1. FLAG-tagged Mcl-1 was prepared as in **Fig. 7a**, using ubiquitin with all lysines substituted with arginine. Mcl-1



*Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations*

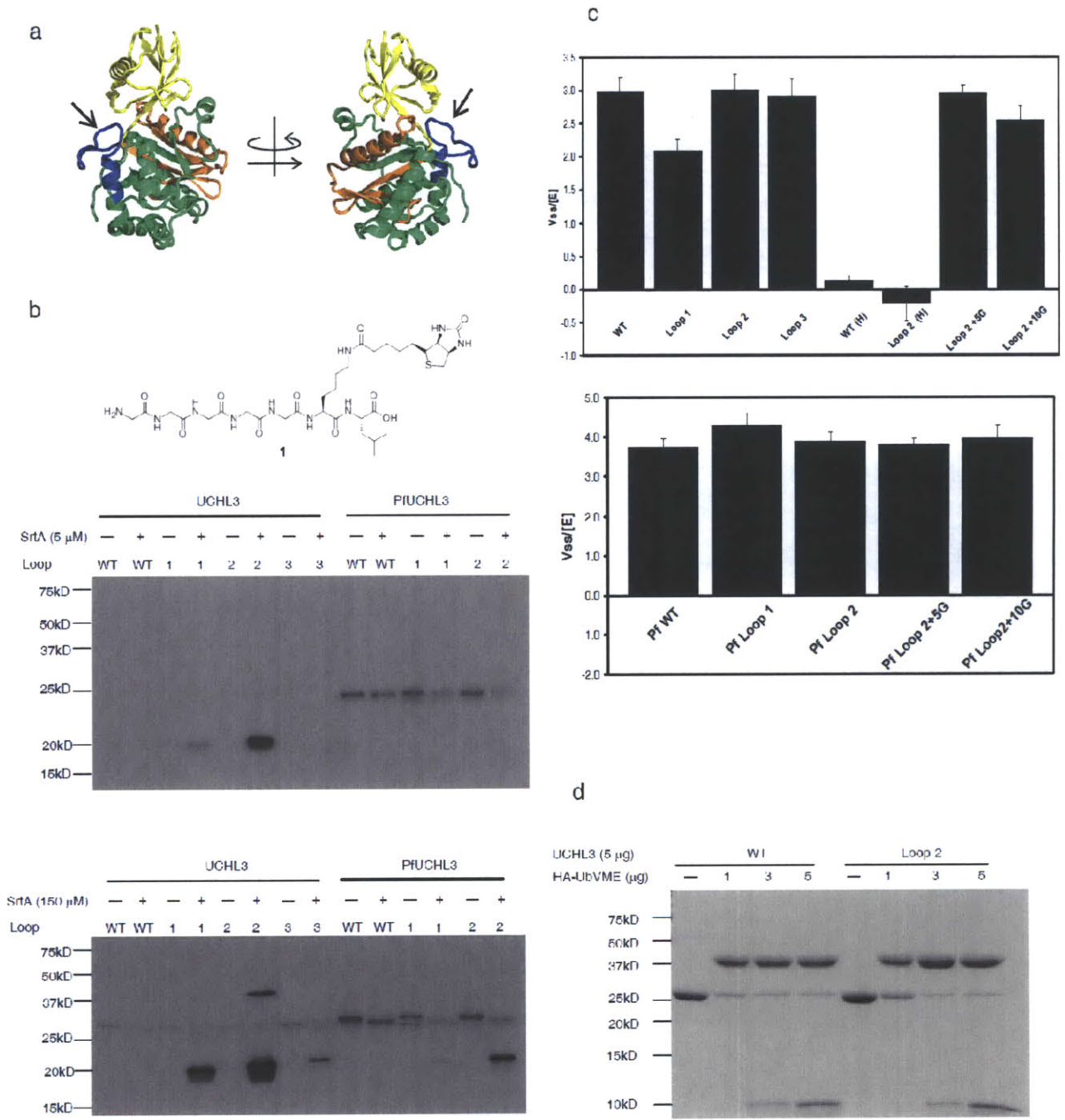
conjugates were digested and analyzed as in **Fig. 7a**.

**Acknowledgements**

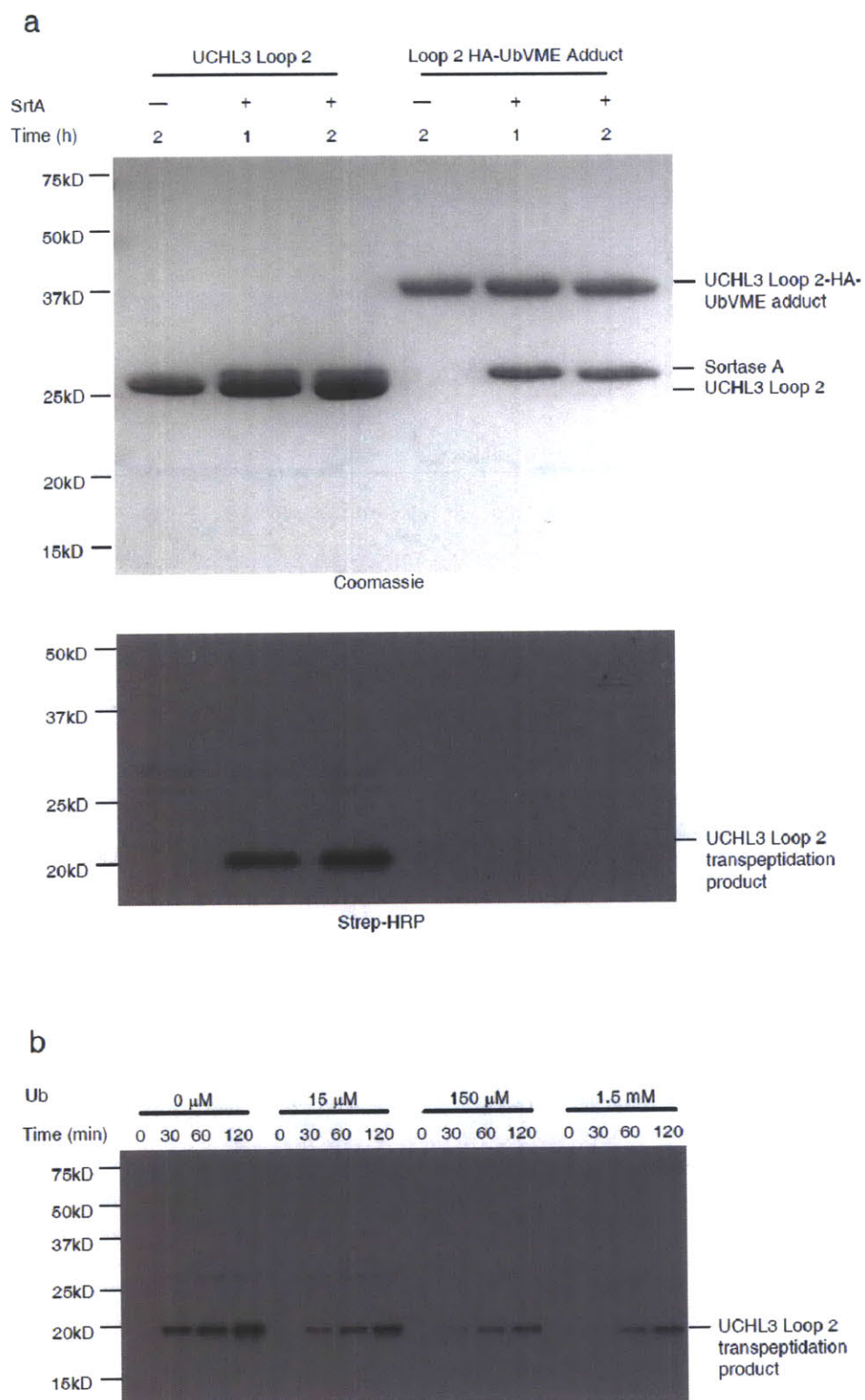
We thank K. Love and R. Sastry for assistance with Mcl-1 ubiquitination reactions and purified FLAG-Mcl-1, C. Schlieker for purified M48<sup>USP</sup>, and G. Grotenbreg for the biotinylated nucleophile peptide.

This research was supported by grants from the NIH.

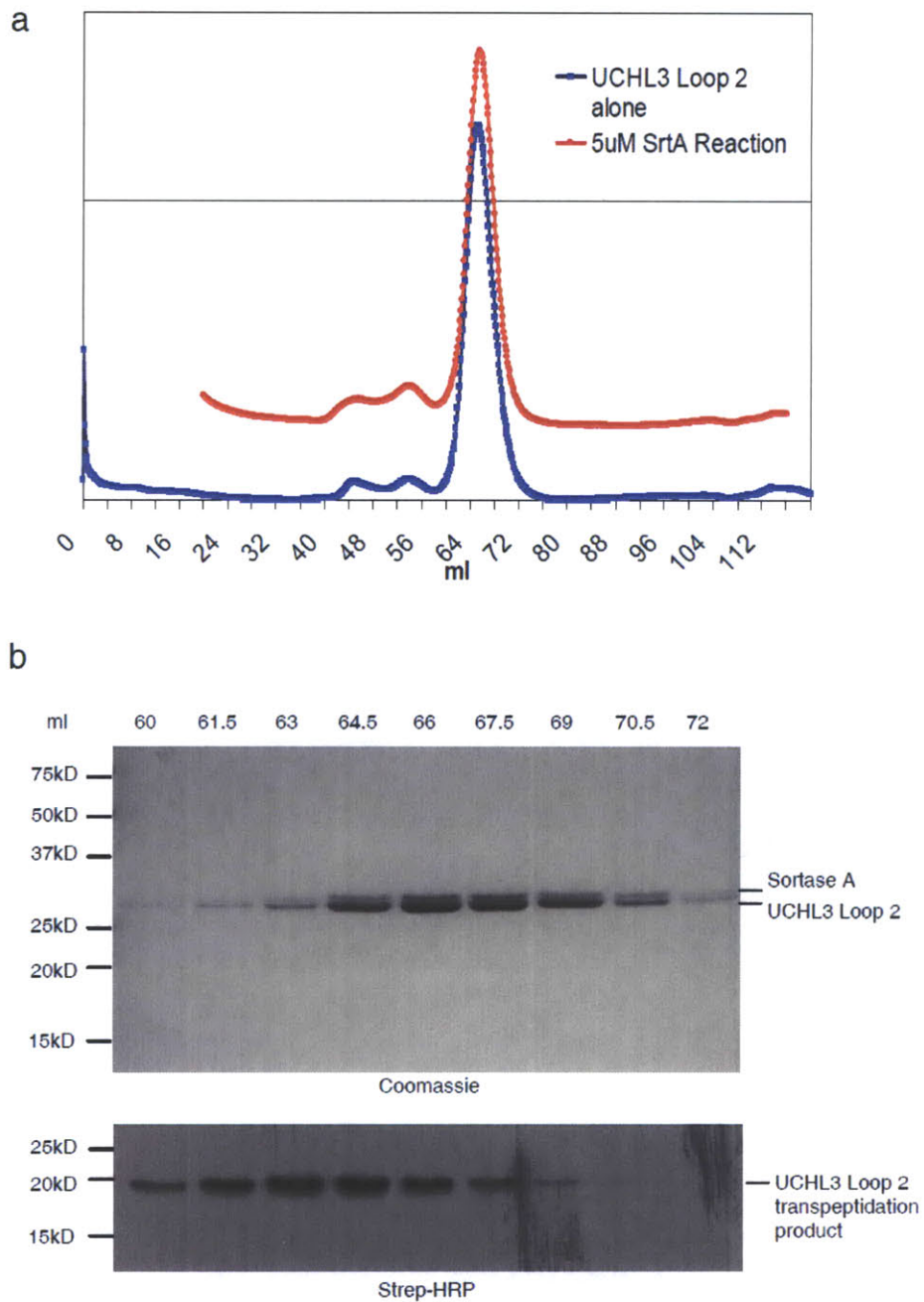
Figure 5.1



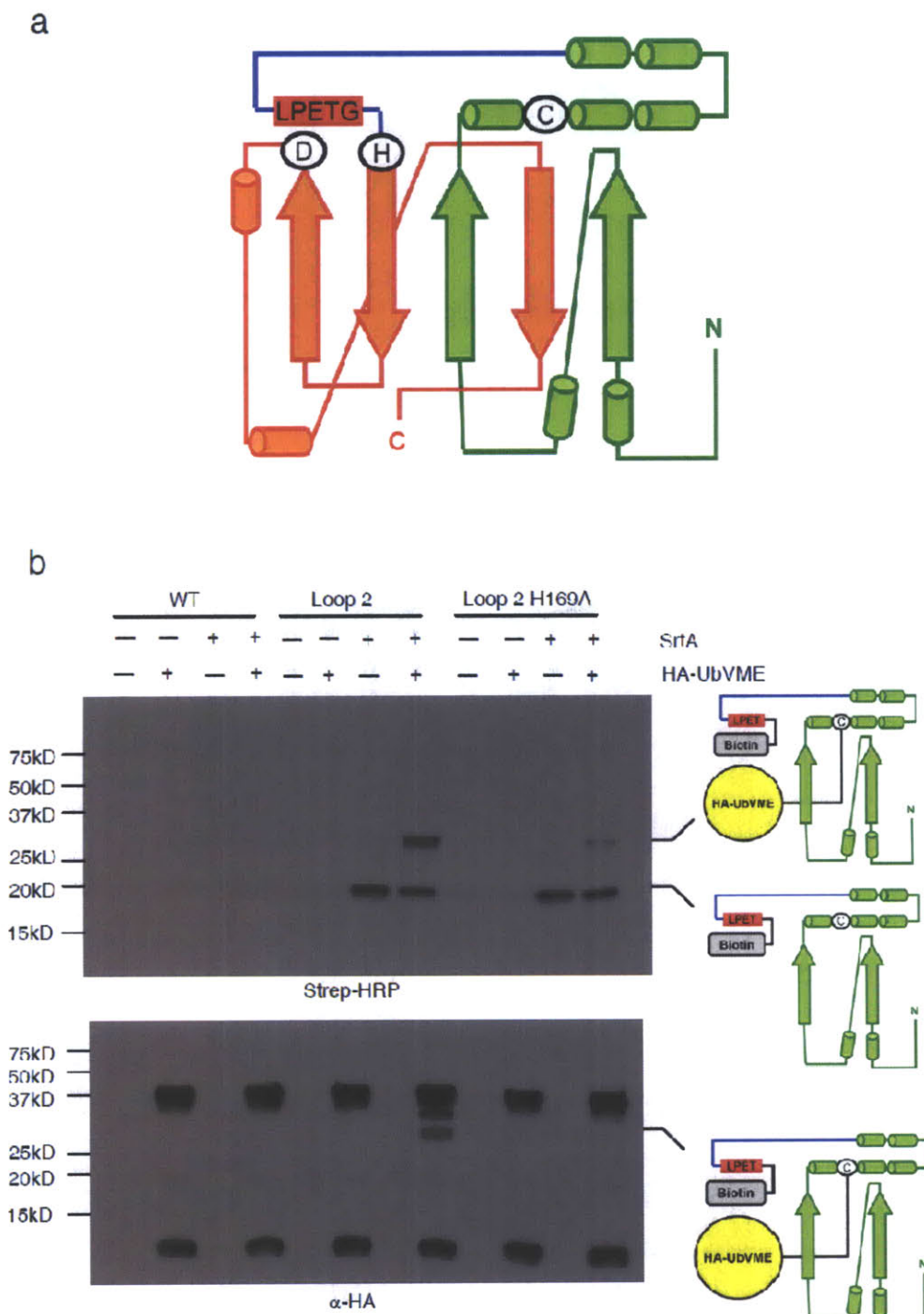
**Figure 5.2**



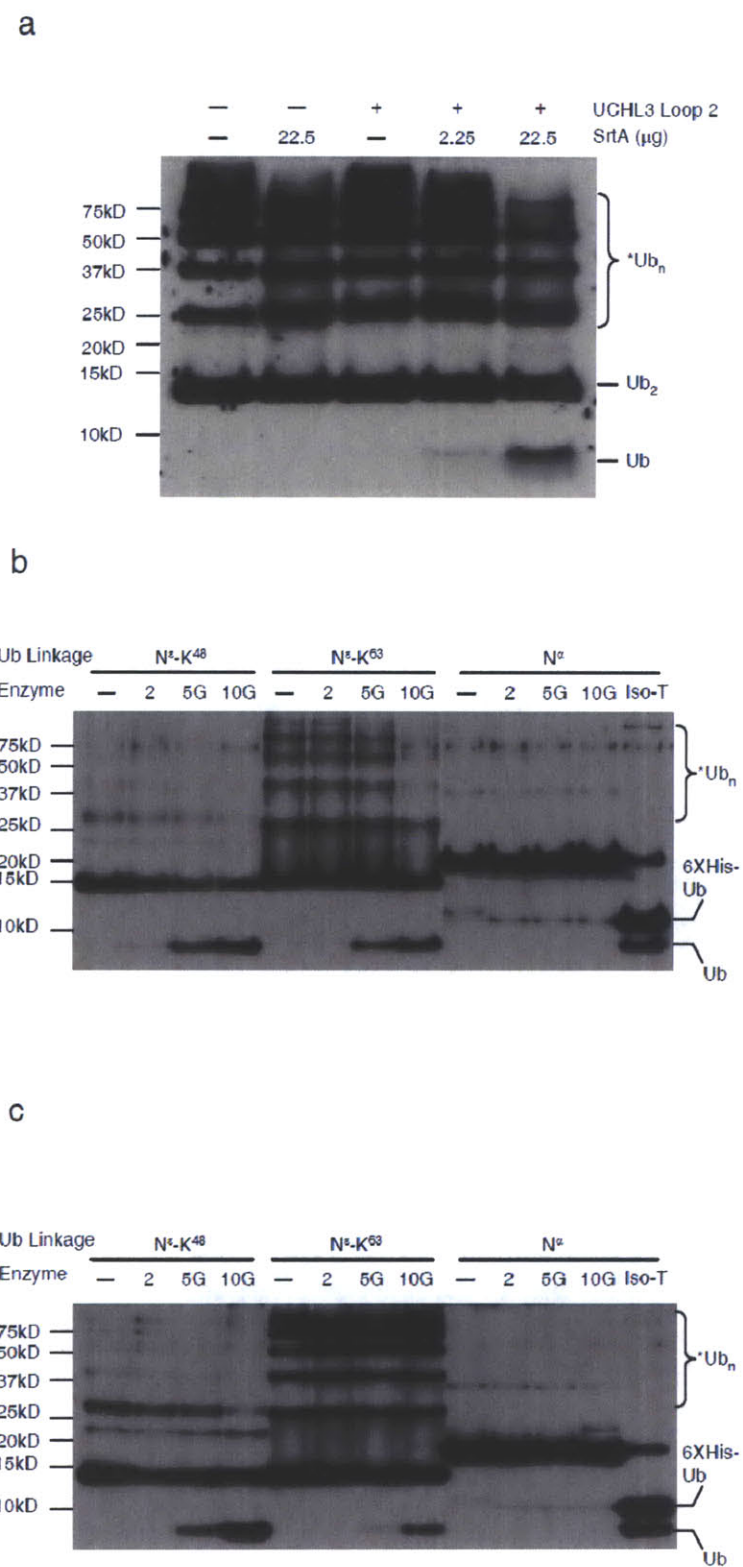
**Figure 5.3**



**Figure 5.4**



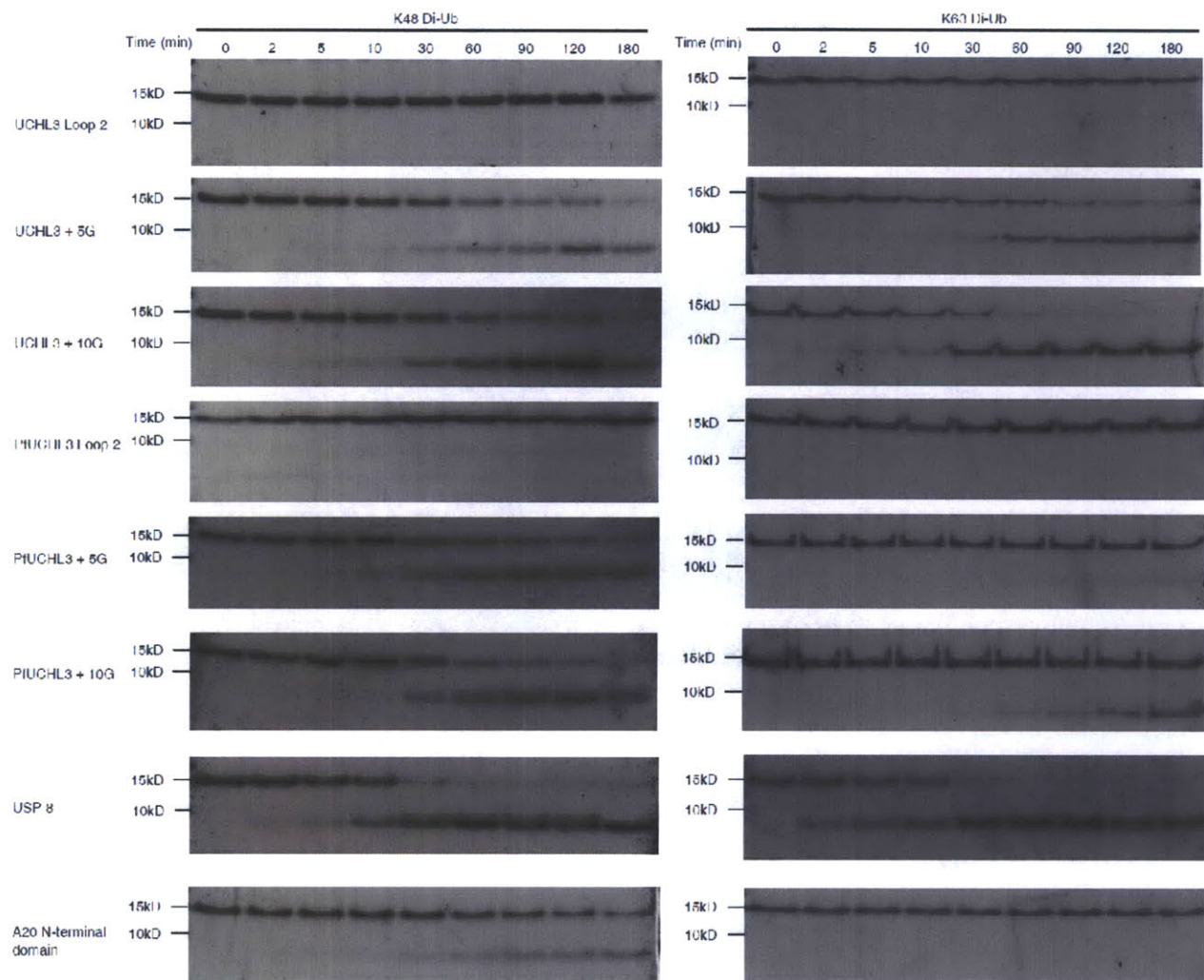
**Figure 5.5**



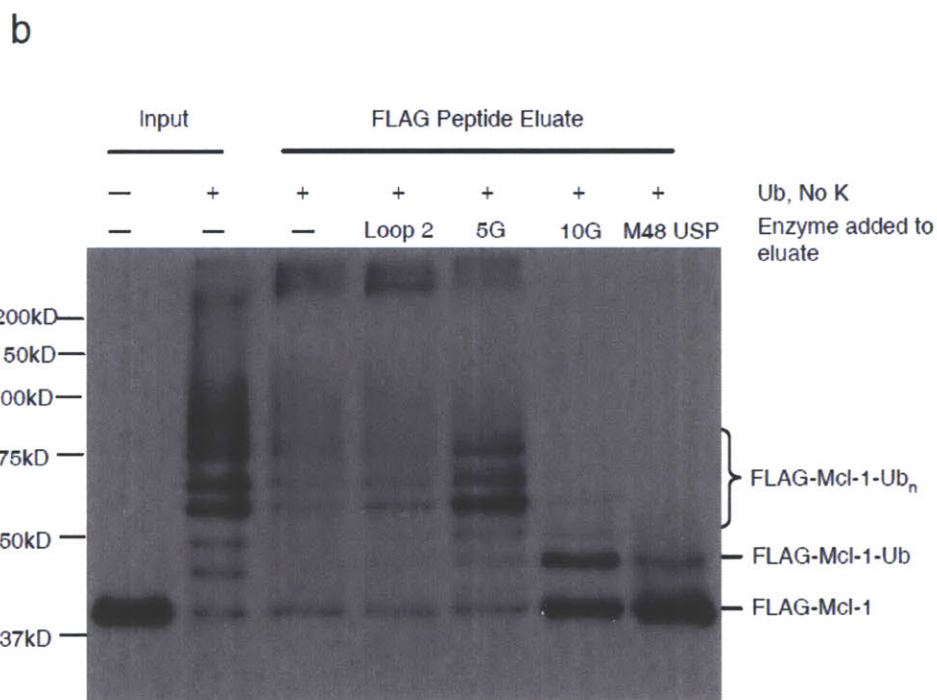
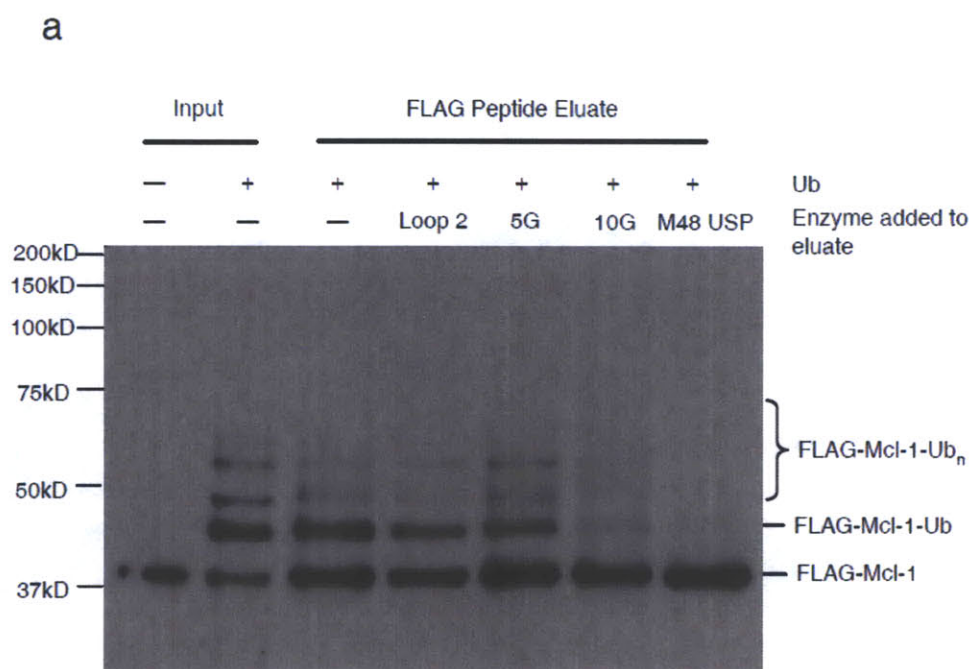


*Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations*

**Figure 5.6**



**Figure 5.7**





Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

**Table 1. K48- diubiquitin hydrolysis rates of engineered UCHL3 and PfUCHL3 enzymes compared to USP8 and the A20 N-terminal domain.**

<i>Enzyme</i>	<i>V/[E]</i> <i>(min<sup>-1</sup>)<sup>a</sup></i>	<i>Rate relative to USP8</i> <i>(%)</i>	<i>Rate relative to A20 NTD</i> <i>(%)</i>
UCHL3+5G	0.281	27.89	195.15
UCHL3+10G	0.410	40.61	284.19
PfUCHL3+5G	0.311	30.80	215.53
PfUCHL3+10G	0.374	37.05	259.22
USP8	1.009	100	700
A20 NTD	0.144	14.27	100

<sup>a</sup> To determine velocities, K48-diubiquitin cleavage reactions were separated by SDS-PAGE (Fig. 6), stained with colloidal coomassie, and the ratio of mono-ubiquitin to total ubiquitin was quantitated. The rate of mono-ubiquitin release versus time was determined by fitting the initial linear data points to a least squares regression line. *NTD*, N-terminal domain.

**Supplementary Figure Legends:**

**Supplementary Figure 5.1. UCHL3 Loop 2 mutant enzyme can be cleaved nearly quantitatively by sortase A to yield the expected transpeptidation product.**

(a) UCHL3 Loop 2 transpeptidation. Sortase A (150  $\mu$ M) and UCHL3 Loop 2 enzyme (50  $\mu$ M) were mixed in the presence of triglycine nucleophile (90 mM) and incubated for the indicated amounts of time at 37 °C. Reactions were analyzed by 12.5% SDS-PAGE and stained with coomassie to visualize total protein.

(b) Reconstructed ESI-MS spectra for UCHL3 Loop 2 transpeptidation reaction.

Reactions from (a) were quenched with 0.1% formic acid and used for liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) analysis. UCHL3 Loop 2 incubated at 37 °C for 8 h (top) and UCHL3 Loop 2 enzyme (50  $\mu$ M) mixed with sortase A (150  $\mu$ M) and 90 mM triglycine nucleophile (bottom). By ESI-MS, the transpeptidation product is the exclusive species of UCHL3 Loop 2 protein seen in the transpeptidation reaction after 8 hours.

**Supplementary Figure 5.2. Head-to-tail linked diubiquitin is likely inaccessible to the UCHL3 active site.**

Structures for K48-linked diubiquitin (a), K63-linked diubiquitin (b), or ISG15 (c) in red were superimposed on the known UCHL3-UbVME (gray) structure by aligning the distal ubiquitin moiety with UbVME. A close up view of the active site (right) is shown, with the catalytic cysteine in blue.

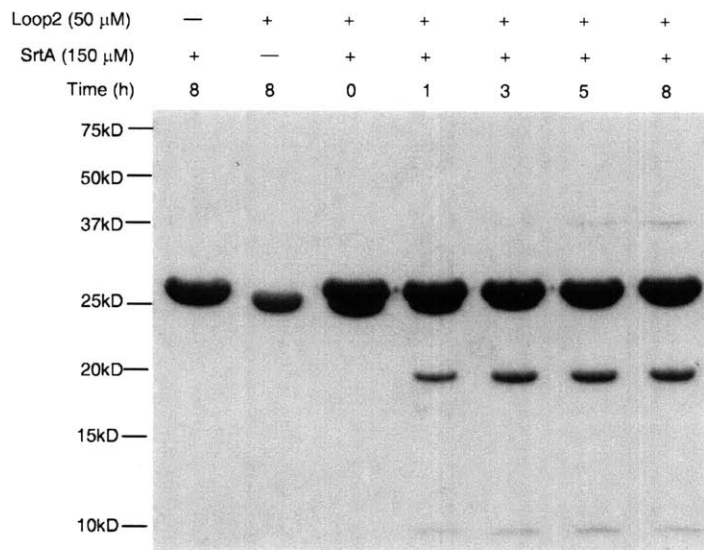
*Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations*

**Supplementary Table 5.1. Table showing the positions of LPETG substitutions in UCHL3 and PfUCHL3.**

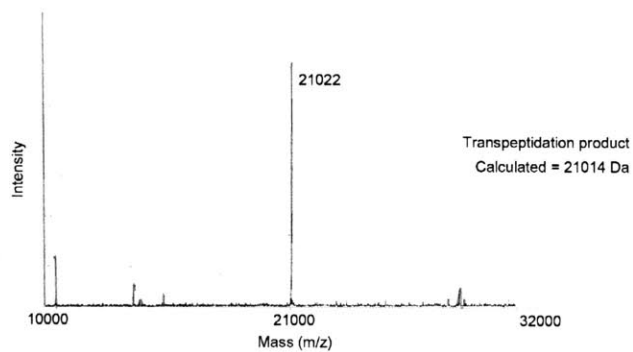
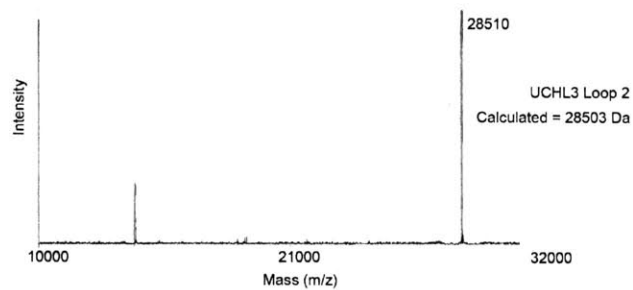
Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations

**Supplementary Figure 5.1.**

**a**

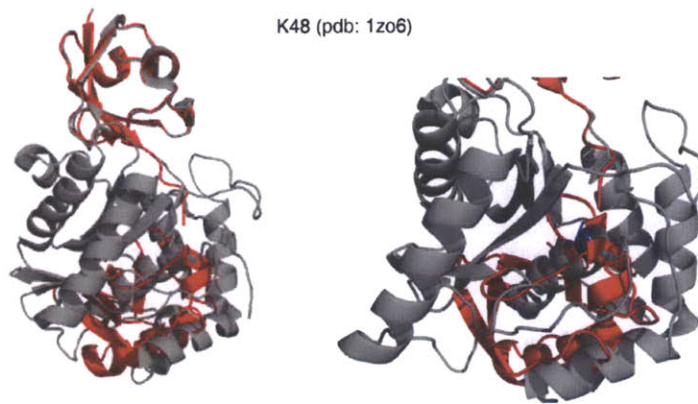


**b**

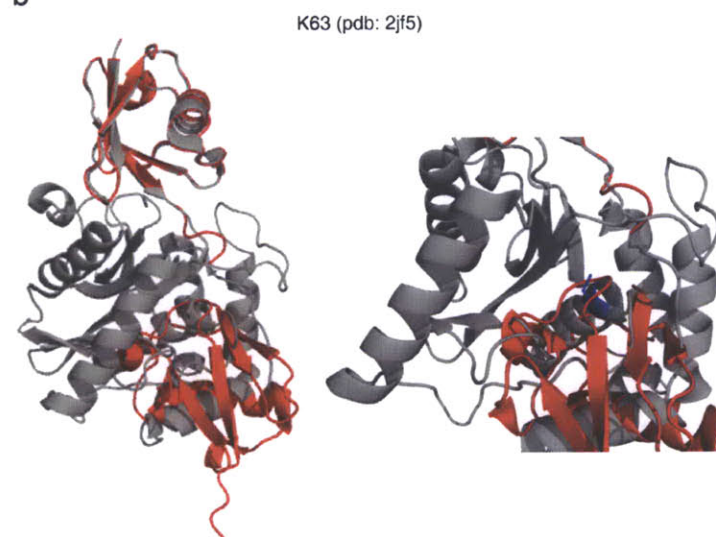


**Supplementary Figure 5.2.**

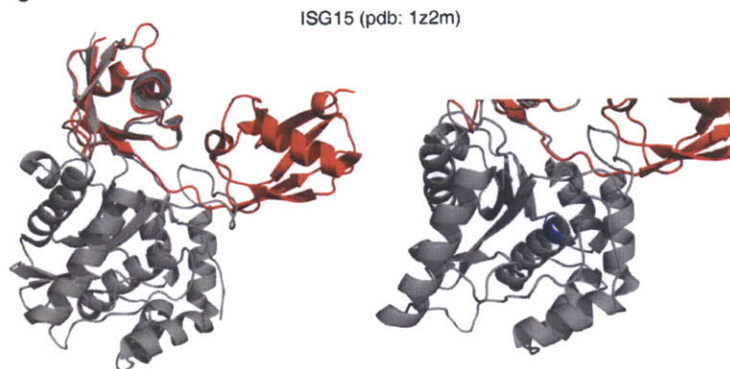
**a**



**b**



**c**



**Supplementary Table 5.1. Table showing the positions of LPETG substitutions in UCHL3 and PfUCHL3.**

<b>Protein</b>	<b>Residue Numbers</b>	<b>WT Sequence</b>	<b>Replaced Sequence</b>
UCHL3 Loop 1	156-160	HEG <b>QTE</b> APSID	HEGL <b>PETG</b> SID
UCHL3 Loop 2	159-163	QTE <b>AP</b> SIDEKV	QTE <b>LPETG</b> EKV
UCHL3 Loop 3	162-166	AP <b>SIDE</b> KVDLH	AP <b>SLPETG</b> DLH
UCHL3 Loop 2 +5G	159-163	QTE <b>AP</b> SIDEKV	QTE <b>LPETGGGGGG</b> EKV
UCHL3 Loop 2 +10G	159-163	QTE <b>AP</b> SIDEKV	QTE <b>LPETGGGGGGGGGGGG</b> EKV
PfUCHL3 Loop 1	151-155	FCG <b>Q</b> VENRDDI	FCGL <b>PETG</b> DDI
PfUCHL3 Loop 2	155-159	VEN <b>R</b> DDILDVD	VEN <b>L</b> PETGDVD
PfUCHL3 Loop 2 +5G	155-159	VEN <b>R</b> DDILDVD	VEN <b>L</b> PETGGGGGGD <b>D</b>
PfUCHL3 Loop 2 +10G	155-159	VEN <b>R</b> DDILDVD	VEN <b>L</b> PETGGGGGGGGGGGGD <b>D</b>

Print Article (reprinted with permission from The Journal of Biological Chemistry):

Supplemental Material can be found at:  
<http://www.jbc.org/content/suppl/2009/12/09/M807172200.DC1.html>

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 6, PP. 3593–3602, FEBRUARY 6, 2009  
© 2009 BY THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, INC. PRINTED IN THE USA.

## Substrate Filtering by the Active Site Crossover Loop in UCHL3 Revealed by Sortagging and Gain-of-function Mutations<sup>\*§</sup>

Received for publication, September 16, 2008, and in revised form, November 24, 2008. Published, JBC Papers in Press, December 1, 2008. DOI: 10.1074/jbc.M807172200

Maximilian W. Popp<sup>†§</sup>, Katerina Artavanis-Tsakonas<sup>‡</sup>, and Hidde L. Ploegh<sup>†§1</sup>

From the <sup>‡</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142 and the <sup>§</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

Determining how deubiquitinating enzymes discriminate between ubiquitin-conjugated substrates is critical to understand their function. Through application of a novel protein cleavage and tagging technique, sortagging, we show that human UCHL3 and the *Plasmodium falciparum* homologue, members of the ubiquitin C-terminal hydrolase family, use a unique active site crossover loop to restrict access of bulky ubiquitin adducts to the active site. Although it provides connectivity for critical active site residues in UCHL3, physical integrity of the crossover loop is dispensable for catalysis. By enlarging the active site crossover loop, we have constructed gain-of-function mutants that can accept substrates that the parent enzyme cannot, including ubiquitin chains of various linkages.

Covalent post-translational modification of proteins with the 76-amino acid ubiquitin (Ub)<sup>2</sup> molecule controls many cellular processes, including protein turnover, trafficking, and transcriptional regulation (1). Ubiquitin conjugation to the ε-amine of lysine residues in target proteins is controlled by a series of enzymes: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzymes (E2), and ubiquitin ligases (E3). Ubiquitin may also be condensed to other ubiquitin molecules, predominantly via internal lysines at positions 48 (K48) or 63 (K63) to form ubiquitin chains. The ubiquitin protein is generated as a head-to-tail fusion from the Ubb and Ubc loci and as a linear fusion to the ribosomal protein, CEP52. Like other regulatory modifications, ubiquitination is reversible. Removal of ubiquitin is the purview of deubiquitinating enzymes (DUBs), comprised of five groups: JAMM motif proteases, ovarian tumor proteases (OTUs), ubiquitin-specific protease (USPs), Machado-Joseph

disease protein domain proteases (MJDs), and ubiquitin C-terminal hydrolases (UCHs) (2, 3). Substrate recognition by these proteases is not well understood and it is highly likely that domains outside of the minimal catalytic unit regulate it. Save for members of the UCH class, no other DUBs have been crystallized in their full-length form. The UCH enzymes are proteins of modest size, capable of hydrolyzing ubiquitin adducts with small leaving groups, and contribute to homeostasis of ubiquitin levels in the cell. It is widely held that many members of this class of ubiquitin-specific hydrolases are unlikely to be involved in editing of ubiquitin-modified proteins, but rather recycle ubiquitin that has been consumed by reactions with small molecules (3). Accordingly, large N-terminal ubiquitin fusion proteins are generally poor substrates for UCH proteases *in vitro* (4). Because detailed structural data are available for several members of the UCH class of DUBs in their full-length form, we chose to study substrate recognition by these enzymes.

The structures of the yeast UCHL3 homologue YUH-1 (5), as well as those of the mammalian UCHL1 (6) and UCHL3 enzymes are known, for UCHL3 both in its free (7) and substrate-occupied form (8). UCHL3 is an unusual enzyme from a topological perspective: it possesses a highly knotted structure, possibly an evolutionary solution to survival in the proteolytic environment of the ubiquitin-proteasome system (9). A distinguishing feature of the enzyme's architecture is the presence of an active site crossover loop that embraces the C-terminal segment of the ubiquitin suicide substrate with which the enzyme was co-crystallized (8). In the absence of substrate, the crossover loop is flexible and not visible in the x-ray structure. The role of this loop, and the relevance of its movements in the course of catalysis is unclear, but it has been proposed that this loop aids in the proper positioning of the substrate, a ubiquitin adduct, in the enzyme active site. In contrast to UCHL3, analysis of the UCHL1 crystal structure (51% sequence identity to UCHL3) reveals occlusion of the active site by a crossover loop that is ordered also in the absence of ubiquitin (6), but perhaps this is because of crystal packing interactions. The comparison of the UCHL1 and UCHL3 enzymes thus leaves the role of the crossover loop in catalysis or positioning of the substrate unresolved.

We engineered a cleavage site in the crossover loop of UCHL3, to explore its contribution to both structure and function. We chose to install a sortase recognition site, because it allows a site-specific cleavage and trans-acylation reaction with concomitant installation of a functionality (biotin, fluorophores) at the site of sortase cleavage (10). By applying the

\* This work was supported, in whole or in part, by grants from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1 and S2 and Table S1.

<sup>1</sup> To whom correspondence should be addressed. E-mail: ploegh@w.mit.edu.

<sup>2</sup> The abbreviations used are: Ub, ubiquitin; Ubl, ubiquitin-like molecule; UCH, ubiquitin C-terminal hydrolase; DUB, deubiquitinating enzyme; USP, ubiquitin-specific protease; Ub-AMC, Ub C-terminal 7-amido-4-methylcoumarin; HA-UbVME, hemagglutinin epitope-tagged ubiquitin vinylmethyl ester; HRP, horseradish peroxidase; LC-ESI-MS, liquid chromatography-electrospray ionization-mass spectrometry; PBS, phosphate-buffered saline; M48, murine cytomegalovirus M48; PUCHL3, *P. falciparum* UCHL3; DTT, dithiothreitol; SrtA, sortase A; Iso-T, Isopeptidase-T (USP5); Mcl-1, myeloid cell leukemia-1 protein.

### Substrate Filtering in UCHL3

sortagging technique, we can simultaneously interrupt the connectivity of a protein peptide backbone and install a tag to track only the cleaved species. Thus, both native and cleaved sortase substrates can be tracked simultaneously in the same reaction mixture. The properties of sortagged UCHL3 inspired us to introduce yet other alterations in the crossover loop, resulting in gain-of-function mutants of UCHL3.

#### EXPERIMENTAL PROCEDURES

**Reagents**—Probe 1 was synthesized by standard *N*-(9-fluorenylmethoxycarbonyl) (Fmoc)-based solid phase peptide chemistry as described (10). Triglycine was purchased from Sigma. Antibodies were purchased from the following vendors: anti-hemagglutinin (HA) tag (antibody 3F10-HRP; horseradish peroxidase), Roche Applied Science; streptavidin-HRP, Amersham Biosciences; anti-FLAG tag M2 (Sigma); anti-ubiquitin (rabbit, Sigma); anti-Mcl-1 (Sigma); goat anti-rabbit-HRP (Southern Biotech). FLAG peptide was purchased from Sigma. Ub C-terminal 7-amido-4-methylcoumarin (Ub-AMC), ubiquitin lacking all lysines, K63-linked diubiquitin, K48-linked diubiquitin, UBEL1, USP8, and A20 N-terminal domain were purchased from Boston Biochem. Purified ubiquitin was purchased from Sigma. Hemagglutinin epitope-tagged ubiquitin vinylmethyl ester (HA-UbVME) was generated as described previously (11).

**Mass Spectrometry**—Liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) was performed on a Micromass LCT mass spectrometer (Micromass MS Technologies) and a Paradigm MG4 HPLC system equipped with a HTC PAL autosampler (Michrom BioResources) and a Waters Symmetry 5  $\mu$ m C8 column (2.1  $\times$  50 mm, MeCN:H<sub>2</sub>O (0.1% formic acid) gradient mobile phase, 150  $\mu$ l/min).

**Cloning and Protein Expression**—Sortase A was expressed and purified as described (12). Murine cytomegalovirus M48 deubiquitinating enzyme domain (M48<sup>USP</sup>) was cloned and purified as described previously (13). The C-terminal residues of the E3 ligase, ARF-BP1 (amino acids 4012–4374), UCHL3, and PfUCHL3 were cloned into pET28a+ (Novagen) with an N-terminal hexahistidine tag. The N-terminal residues of Mcl-1 (myeloid cell leukemia-1 protein, amino acids 1–327) were cloned as an N-terminal FLAG peptide fusion into the vector pET16b (Novagen) carrying an N-terminal His<sub>10</sub> tag. The final Mcl-1 construct consists of an N-terminal His<sub>10</sub> tag followed by a FLAG tag and Mcl-1<sub>1–327</sub>. UCHL3 point mutations and LPETG substitutions were generated by site-directed mutagenesis using a QuikChange kit (Stratagene). Loop expanded versions of UCHL3 and PfUCHL3 were generated by inserting 5 or 10 glycine residues immediately after the LPETG sequence. Hexahistidine-tagged linear diubiquitin was also cloned into pET28a+. Proteins were expressed in BL-21 *E. coli* or Rosetta (FLAG-Mcl-1) DE3(PlysS) cells (Novagen), resuspended in lysis buffer (50 mM Tris, 150 mM NaCl, 10 mM imidazole, 10% glycerol, pH 7.2) and lysed by French press. Proteins were purified from clarified lysates with nickel-nitrilotriacetic acid agarose (Ni-NTA, Qiagen) and eluted in lysis buffer supplemented with 500 mM imidazole. Eluted fractions were dialyzed extensively to remove imidazole. Protein concentrations were determined by the Bradford method (Bio-Rad).

**Immunoblotting**—Proteins were separated by Tris/Glycine SDS-PAGE or Tris/Tricine PAGE (Ub chain digestion experiments) and transferred to nitrocellulose membranes or polyvinylidene fluoride (ubiquitin chain digestion experiments). Membranes were blocked with 5% nonfat dried milk in phosphate-buffered saline supplemented with Tween 20 (PBS, 0.1% Tween 20, pH 7.4) overnight at 4 °C or for 1 h at room temperature. Membranes were washed with PBST and incubated with the indicated antibodies for 1 h. Streptavidin-HRP and anti-HA (3F10-HRP) blots were then washed and developed with Western Lighting Chemiluminescence Reagent Plus (PerkinElmer Life Sciences). All other blots were washed, incubated with a goat anti-rabbit-HRP conjugate, and developed.

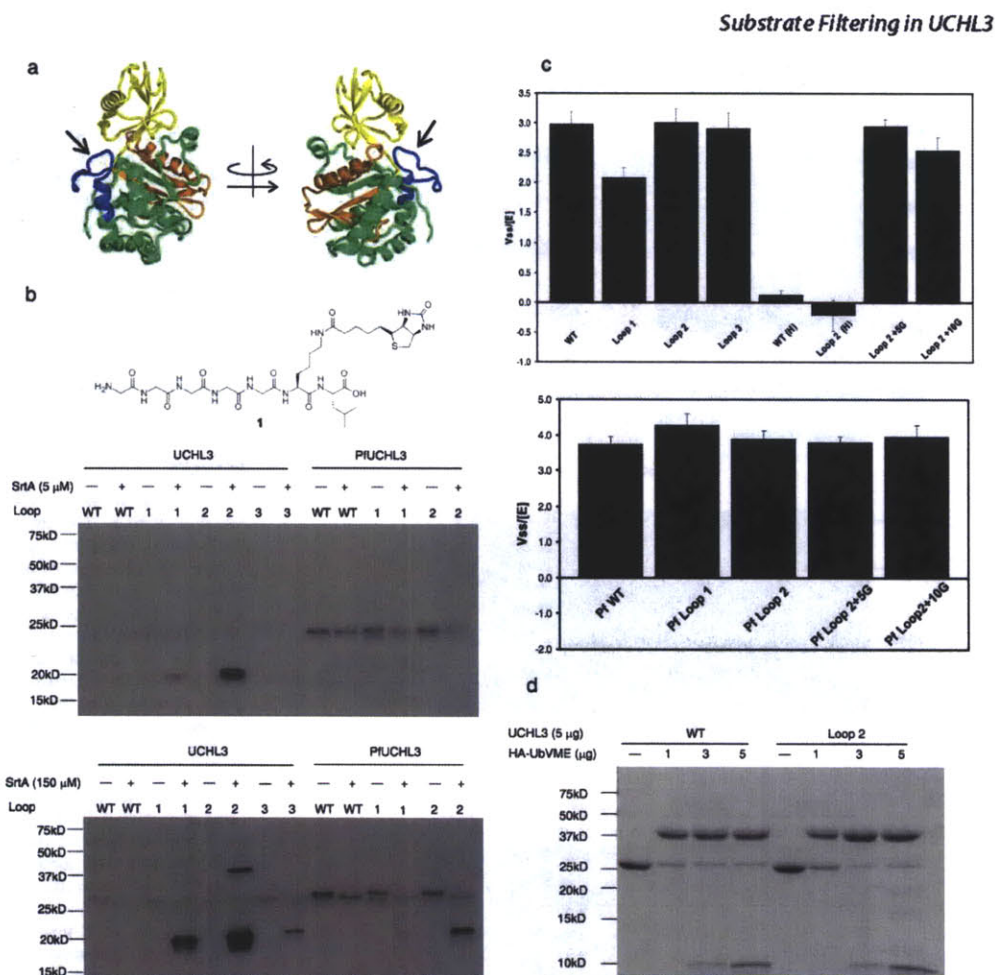
**Sortase Cleavage and Gel Filtration of Sortagging Reactions**—Sortagging of UCHL3 (10  $\mu$ M) was performed by incubating with the indicated concentrations of sortase A (Srt A) in Srt buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM CaCl<sub>2</sub>), and 5 mM probe 1 in a 25- $\mu$ l volume. Reactions were incubated at 37 °C for 2 h and halted with sample buffer. For gel filtration, a 1-ml reaction (5  $\mu$ M sortase) was purified by size exclusion chromatography on S75 Sephadex resin using 20 mM Tris, 50 mM NaCl, pH 8.0 buffer as eluent. Fractions were collected and analyzed by SDS-PAGE.

**Kinetic Measurements**—Ub-AMC assays were performed in assay buffer (50 mM Tris/HCl, 150 mM NaCl, 2 mM EDTA, 2 mM DTT, 1 mg/ml bovine serum albumin, pH 7.5) at 25 °C. Enzyme concentrations were determined by Bradford assay (Bio-Rad). UCHL3 mediated Ub-AMC hydrolysis was performed in a total volume of 30  $\mu$ l with 10 pM UCHL3 mutants and 62.5 nM Ub-AMC in a 384-well NUNC black plate. PfUCHL3 assays were performed with 25 pM enzyme and 125 nM UbAMC. Data were collected with a Spectramax M2 plate reader (Molecular Devices) with a 368 nm/467 nm filter pair and a 455-nm cutoff. Velocities were determined by fitting the initial linear data points to a least-squares regression line.

**HA-UbVME Labeling and Adduct Purification**—Sortase cleavage reactions (5  $\mu$ M sortase) of the indicated UCHL3 mutants were performed for 2 h at 37 °C and 5  $\mu$ l was diluted with 7  $\mu$ l of labeling buffer (20 mM Tris pH 8.0, 150 mM NaCl), DTT (1 mM final concentration) and 1.5  $\mu$ g of HA-UbVME was added, and reactions were incubated at 37 °C for 1 h. Reactions were halted with sample buffer and loaded onto 12.5% SDS-PAGE for analysis. For HA-UbVME adduct purification, a large-scale reaction (1 ml) was prepared, omitting sortase, and purified by ion-exchange chromatography as described (8).

**Ubiquitin Polymer Digestion**—K63-linked diubiquitin cleavage by sortagged UCHL3 Loop 2 enzyme was performed by incubating 7  $\mu$ g of UCHL3 Loop 2 with either 2.25  $\mu$ g or 22.5  $\mu$ g of SrtA and 5 mM Probe 1 in Srt buffer for 2 h at 37 °C. K63-linked diubiquitin (0.5  $\mu$ g) was then added as well as 1 mM DTT and incubated for 1 h at 37 °C. Reactions were halted with Tris/Tricine sample buffer and subjected to 12.5% Tris/Tricine PAGE followed by anti-ubiquitin immunoblot. Digestion of ubiquitin chains by UCHL3 loop extension mutants was performed in Srt buffer with 180 ng of each enzyme, 1 mM DTT, and 1  $\mu$ g of the indicated ubiquitin chain in a total volume of 7  $\mu$ l. Reactions were halted with Tris/Tricine sample buffer and loaded onto a 12.5% Tris/Tricine gel for anti-ubiquitin immu-





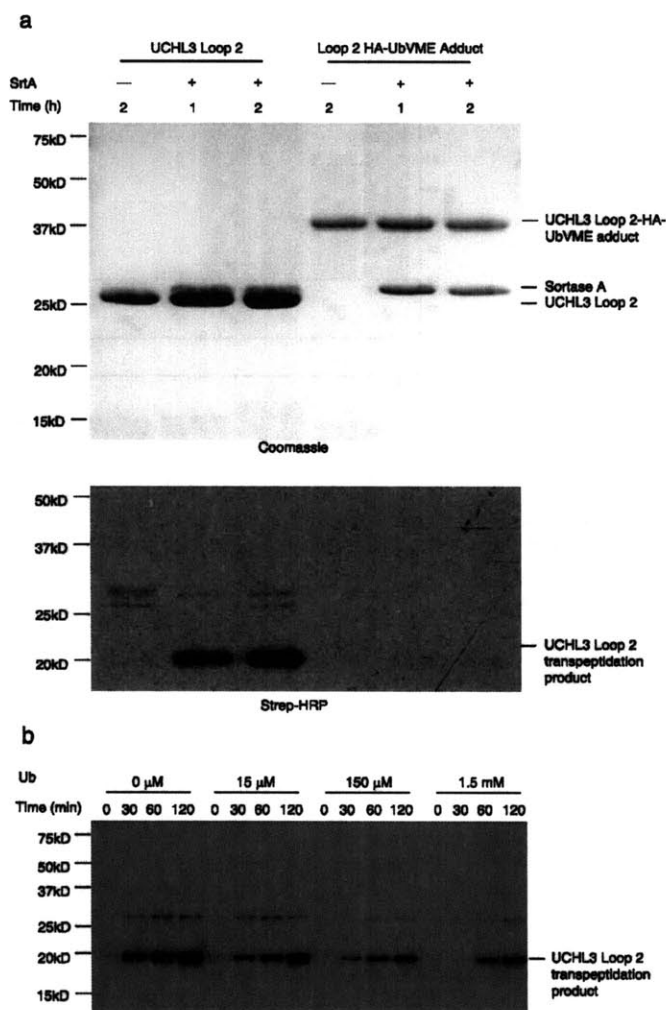
**FIGURE 1. Characterization of UCHL3 and PFIUCHL3 sortase substrates.** *a*, crystal structure of UCHL3 in complex with UbVME (PDB: 1xd3). UbVME is in yellow with the crossover loop (blue) spanning the two halves of UCHL3 (green and orange). The position of the Loop 2 substitution is indicated by an arrow. *b*, LPETG substitution in the crossover loop renders UCHL3 and PFIUCHL3 susceptible to sortase-mediated transpeptidation. Wild-type and LPETG substituted UCHL3 and PFIUCHL3 mutants (10  $\mu$ M) were exposed to sortase (5  $\mu$ M, top blot; 150  $\mu$ M, bottom blot) and biotinylated oligoglycine probe **1** (5 mM) and analyzed by streptavidin-HRP immunoblot to detect the  $\sim$ 20 kDa transpeptidation product. *c*, LEPTG-substituted UCHL3 and PFIUCHL3 mutants catalyze Ub-AMC hydrolysis. Ub-AMC hydrolysis by wild-type and mutant UCHL3 (10 pM) was measured using saturating concentrations of Ub-AMC (62.5 nM). Data are plotted as the ratio of velocity to the enzyme concentration, with  $n \geq 3$  (top). Enzymes with an H169A mutation are indicated by an (H), and enzymes with 5 or 10 glycines added to the crossover loop directly after the LPETG sequence are indicated by +5G and +10G. Ub-AMC hydrolysis (125 nM) was also measured for wild-type and mutant PFIUCHL3 enzymes (25 pM). Data are plotted as the ratio of the velocity to the enzyme concentration, with  $n = 3$  (bottom). *d*, an LPETG substituted UCHL3 mutant is labeled by a ubiquitin suicide substrate. HA-UbVME was titrated into a fixed amount (5  $\mu$ g) of wild-type UCHL3 or the human UCHL3 Loop 2 mutant enzyme. Reactions were analyzed by 12.5% SDS-PAGE and stained with Coomassie Blue to visualize total protein.

nolot analysis. For kinetic measurements, 500 nM of enzyme was incubated with 10  $\mu$ M of either K63- or K48- linked diubiquitin in buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 2 mM EDTA, 2 mM DTT). At the indicated times, samples were withdrawn, quenched with Tris/Tricine loading buffer and subjected to 10% Tris/Tricine PAGE. Gels were stained with colloidal Coomassie and bands for K48 diubiquitin hydrolysis were quanti-

tated with ImageJ. Velocities were determined by fitting the initial linear data points to a least squares regression line. Higher order K63-linked ubiquitin conjugates visible in Colloidal Blue-stained gels precluded quantitation of K63 diubiquitin hydrolysis.

**Mcl-1 Ubiquitination and Digestion**—FLAG-Mcl-1 (1  $\mu$ g) was incubated with 100 ng of human UBE1 (E1), 1  $\mu$ g of UbC<sup>H7</sup>

**Substrate Filtering in UCHL3**



**FIGURE 2. Ubiquitin binding alters susceptibility to sortase-mediated transpeptidation.** *a*, UCHL3-HA-UbVME adduct is refractory to sortase-mediated transpeptidation. Free UCHL3 Loop 2 enzyme (10  $\mu$ M) and the purified HA-UbVME adduct (10  $\mu$ M) were exposed to SrtA (5  $\mu$ M) and biotinylated probe 1 (5 mM) for the indicated times. Reactions were analyzed by 12.5% SDS-PAGE and either stained with Coomassie Blue (top) or transferred to nitrocellulose for streptavidin-HRP immunoblot (bottom). *b*, ubiquitin binding competes with sortase-mediated transpeptidation. UCHL3 Loop 2 enzyme was incubated with SrtA (5  $\mu$ M) and probe 1 (5 mM) for the indicated amounts of time in the absence or presence of various concentrations of purified ubiquitin. The transpeptidation product was detected by streptavidin-HRP immunoblot.

(E2), 10  $\mu$ g of Arf/BP1 (E3), and 100  $\mu$ g of Ub with an ATP-regenerating system (14) for 90 min at room temperature.<sup>3</sup> Reactions were quenched with NET buffer (50 mM Tris-HCl,

<sup>3</sup> K. R. Love, R. K. Sastry, E. Spooner, and H. L. Ploegh, manuscript in preparation.

pH 7.4, 0.5% Nonidet P-40, 150 mM NaCl, 5 mM EDTA) and incubated overnight at 4 °C with anti-FLAG M2 antibody. Immune complexes were recovered with protein G beads (Sigma), washed extensively with NET buffer, and eluted by incubating with FLAG peptide (250  $\mu$ g/ml in Srt Buffer) at 25 °C with shaking for 30 min. Eluates were collected, pooled, divided, and incubated with 1 mM DTT and UCHL3 or P $\mu$ UCHL3 mutant enzymes (800 ng) for 2 h at 37 °C. Samples were analyzed by SDS-PAGE (10%), transferred to nitrocellulose, and anti-Mcl-1 immunoblot was performed.

**RESULTS**

*Engineering a Sortase Cleavage Site into UCHL3*—The presence of the unusual active site crossover loop in UCHL3 suggests that it may play a role in substrate selection by making sequence-specific contacts to the substrate: either to the ubiquitin (Ub) or ubiquitin-like (Ubl) moiety or to the attached leaving group. Alternatively, the crossover loop may play a role in stabilization of the catalytic center of the enzyme. The catalytic cysteine (Cys-95) and oxyanion-hole stabilizing glutamine (Gln-89) are separated from the general base histidine (His-169) and aspartic acid (Asp-184) residues by the crossover loop, with His-169 lying only three residues from the C-terminal end of the loop. The crossover loop not only traverses the active site of the enzyme, but also provides connectivity for the two halves of the catalytic center (Figs. 1*a* and 4*a*). The loop thus bridges key residues, bringing them into proximity for catalysis and possibly imparting stability to the active site. Alternatively, the active site crossover loop has been suggested to act as a substrate filter, limiting the size of

the ubiquitin C-terminal leaving group (7), a possibility that has been suggested but never experimentally addressed. To examine these possibilities, we engineered an LPETG sortase cleavage site into the crossover loop of both human UCHL3 and the *Plasmodium falciparum* homolog, P $\mu$ UCHL3. We included P $\mu$ UCHL3 in our analysis because the two enzymes



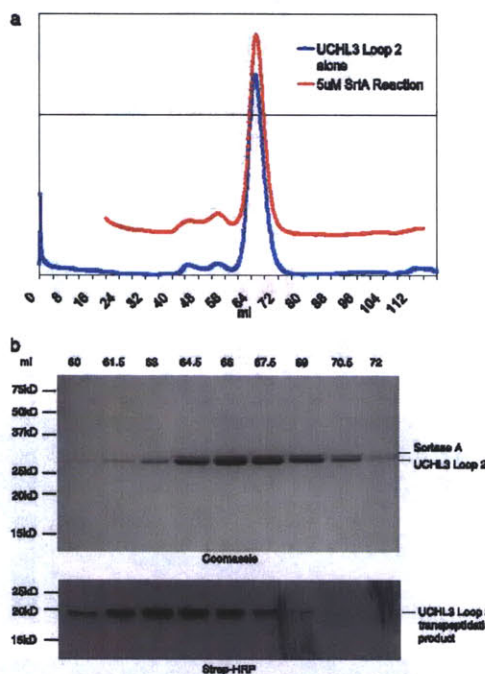
are structurally similar,<sup>4</sup> yet possess completely unrelated sequences in their crossover loop. As will be described below, results obtained for UCHL3 and PfUCHL3 are largely similar and thus allow generalization of our conclusions.

We inserted the LPETG cleavage site at three separate positions in the human UCHL3 (UCHL3) crossover loop and at two positions in the *P. falciparum* UCHL3 (PfUCHL3) crossover loop (supplemental Table S1 and Fig. 1a), and expressed the mutant enzymes in *Escherichia coli*. When exposed to SrtA and the biotinylated oligo-glycine nucleophile 1, we observed successful transacylation for all mutant enzymes, albeit at different efficiencies (Fig. 1b). The loop in both human and plasmodium UCHL3 therefore adopts a flexible conformation in solution. The extent of cleavage of the UCHL3 Loop 2 variant by SrtA can be modulated from 10% to nearly 90% by incubation with increasing amounts of SrtA and increasing time (supplemental Fig. S1). Ubiquitin hydrolase activity was largely unchanged, as assessed by hydrolysis of the ubiquitin-7-amino-4-methylcoumarin (Ub-AMC) substrate (Fig. 1c). In addition, the LPETG substituted human UCHL3 Loop 2 mutant formed a covalent adduct with a hemagglutinin (HA)-tagged ubiquitin vinylmethyl ester suicide substrate (HA-UbVME), designed to react with the active site cysteine in DUBs, with similar efficiency as did wild-type UCHL3 (Fig. 1d). We chose to use the human UCHL3 Loop 2 and PfUCHL3 Loop 2 mutants for further analysis because these mutants retained near wild-type Ub-AMC hydrolysis activity and were most efficiently cleaved by sortase. We conducted subsequent sortagging experiments at low sortase concentrations that yield incomplete conversion to the transpeptidation product to track the properties of both cleaved and uncleaved UCHL3 species under identical conditions.

**Crossover Loop Connectivity Is Dispensable for Ubiquitin Docking**—The crossover loop in liganded UCHL3 likely assumes an ordered conformation in solution (8). When purified UCHL3-HA-UbVME adduct is exposed to SrtA, no transpeptidation is observed (Fig. 2a), consistent with previous observations that the LPETG motif must be placed in a flexible, unstructured region (10). In addition, titration of ubiquitin into the sortase cleavage reaction successfully inhibits transpeptidation (Fig. 2b). Therefore, in solution, the liganded form of UCHL3 possesses a rigid crossover loop refractory to attack by sortase. The highly knotted structure of UCHL3 suggests that it is possible to nick the crossover loop without complete unfolding of the polypeptide. We find only a slight increase in Stokes' radius upon cleavage of the crossover loop (Fig. 3), suggesting that the nicked preparation does not undergo gross alterations in folding state. Nonetheless, the slightly larger hydrodynamic radius of the transpeptidation product is likely due to a more relaxed conformation caused by opening of the flexible crossover loop and installation of the 771-Da biotinylated probe (Fig. 3b). We further examined whether connectivity of the active site is essential for maintenance of UCHL3 structure by first cleaving the active site loop with small amounts of sortase and then delivering an HA-UbVME suicide substrate. Incomplete

<sup>4</sup> K. Artavanis-Tsakonas, W. A. Welhofen, R. Gaudet, and H. L. Ploegh, manuscript in preparation.

### Substrate Filtering in UCHL3

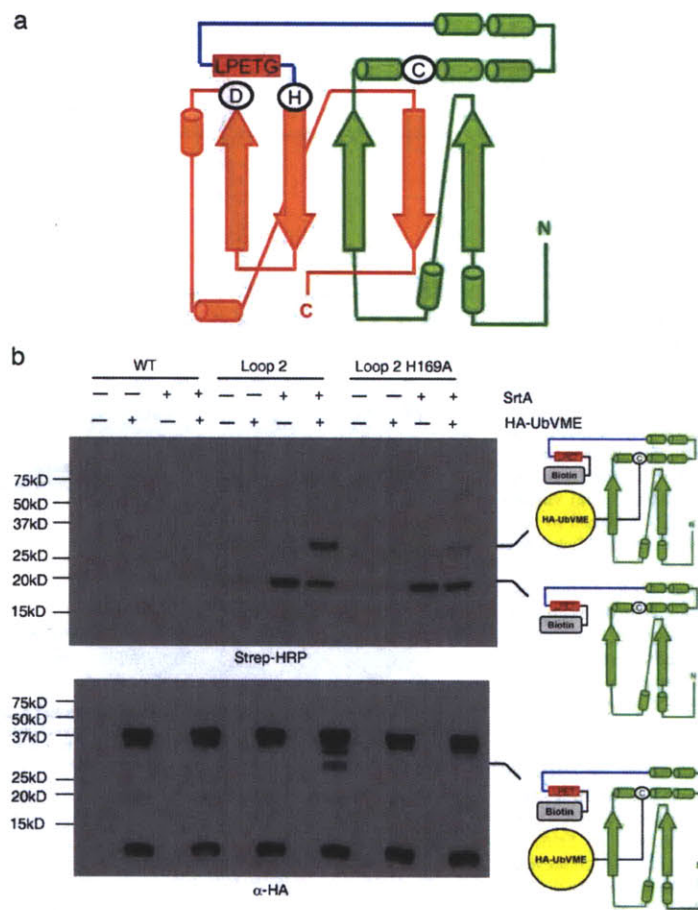


**FIGURE 3. Cleaved and intact UCHL3 migrate with similar Stokes' radius by gel filtration chromatography.** a, similar gel filtration chromatography elution profiles of uncleaved UCHL3 Loop 2 alone (10  $\mu$ M); blue line and UCHL3 Loop 2 (10  $\mu$ M) incubated with SrtA (5  $\mu$ M) and probe 1 (5 mM); red line. Under these conditions, the majority of the UCHL3 population is uncleaved. Reactions were subjected to size exclusion chromatography on a Superdex S-75 column, and the UV absorbance at 280 nm was recorded. b, intact UCHL3 Loop 2 protein, cleaved UCHL3 Loop 2 protein, and sortase A co-migrate by size exclusion chromatography. Fractions corresponding to the major peak (60–72 ml) in the sortase cleavage reaction (red line in a) where the majority of the UCHL3 population is uncleaved were collected and analyzed by 12.5% SDS-PAGE followed by Coomassie Blue staining (top) to detect total protein or streptavidin-HRP immunoblot (bottom) to detect the transpeptidation product.

cleavage by sortase allows the properties of both the cleaved and uncleaved UCHL3 species to be examined simultaneously in one and the same reaction mixture. Indeed, a fraction of the cleaved UCHL3 successfully reacts with HA-UbVME, as shown by the appearance of a species that is both biotinylated and anti HA-reactive (Fig. 4b). Because the HA-UbVME adduct is refractory to cleavage by sortase, sortase-mediated transpeptidation must have preceded HA-UbVME addition. Because of the electrophilicity of the vinyl methylester group, we predicted that HA-UbVME adduct formation would require only the cysteine nucleophile and not the general base histidine residue. We find that mutation of His-169 to alanine results in HA-UbVME reaction levels comparable to wild-type UCHL3 (Fig. 4b, bottom), but eliminates Ub-AMC hydrolysis (Fig. 1c). Reaction of the catalytic cysteine with HA-UbVME requires proper folding of the enzyme to allow ubiquitin recognition, as evidenced by highly specific alkylation of DUB enzymes in whole cell lysate



Substrate Filtering in UCHL3



**FIGURE 4. Crossover loop integrity is not necessary for maintenance of UCHL3 active site structure.** a, schematic showing the UCHL3 secondary structure elements with  $\beta$  sheets represented as arrows and  $\alpha$  helices as cylinders. The position of the crossover loop (blue) is shown relative to the catalytic residues (circled letters). Upon sortase-mediated transpeptidation, probe 1 is affixed to the half shown in green, interrupting the connectivity to the half shown in orange. Note that the half of the enzyme that receives probe 1 contains the catalytic cysteine residue that reacts with HA-UbVME. b, cleaved UCHL3 preparations react with a ubiquitin suicide substrate. Wild-type UCHL3, UCHL3 Loop 2, and UCHL3 Loop 2 H169A mutant enzymes (10  $\mu$ M) were exposed to SrtA (5  $\mu$ M) and probe 1 (5 mM) for 2 h at 37  $^{\circ}$ C and subsequently incubated with HA-UbVME (1.5  $\mu$ g) for 1 h at 37  $^{\circ}$ C. Reactions were analyzed by 12.5% SDS-PAGE followed by streptavidin-HRP immunoblot and anti-HA immunoblot. The identities of the bands detected are shown (right).

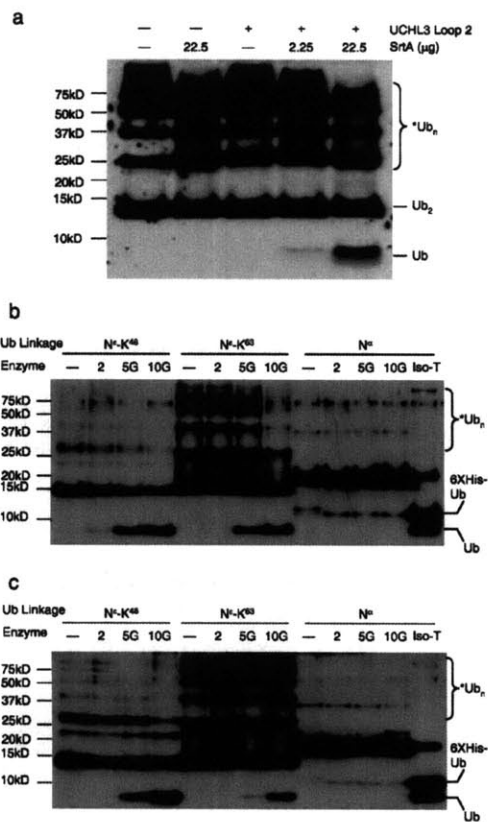
experiments (15). Moreover, DUB protein preparations typically contain a small fraction of unfolded protein that does not react with electrophilic ubiquitin derivatives (16) (Fig. 1*d*), despite the presence of a cysteine residue. Thus HA-UbVME reactivity reflects competency to bind ubiquitin and to position the catalytic cysteine in proximity of the electrophilic trap, and not ubiquitin hydrolase activity *per se*. Taken together, we conclude that connectivity of the active site crossover loop is not

essential for maintenance of UCHL3 structure, as assessed by the ability of the cleaved preparation to interact with ubiquitin. **The Crossover Loop Restricts Substrate Size**—We reasoned that if the role of the active site crossover loop is to restrict the size of the ubiquitin C-terminal leaving group, then ablation of the connectivity should result in relaxed leaving group specificity. Indeed, UCHL3 with a nicked active site crossover loop hydrolyzes K63-linked ubiquitin chains to monomeric ubiquitin, while the uncleaved preparation and sortase itself fail to do so (Fig. 5*a*). Incubation of UCHL3 with increasing concentrations of SrtA results in greater oligo-ubiquitin hydrolysis, indicating that it is the cleaved UCHL3 species that mediates liberation of mono-ubiquitin. This is consistent with the HA-UbVME labeling experiments and indicates that a portion of the cleaved UCHL3 species not only retains proper structure, but also ubiquitin hydrolase activity. The modest ubiquitin hydrolysis activity of the cleaved UCHL3 likely indicates that a fraction of the cleaved material is inactive upon transpeptidation, perhaps because the structure is subtly destabilized. Nevertheless, the gain-of-function associated with the cleaved UCHL3 preparation indicates that at least some fraction is competent to attack polyubiquitin chains.

**Expansion of the Active Site Crossover Loop Results in Gain-of-function Mutants**—Based on these results, we reasoned that insertion of additional amino acids in the active site crossover loop to extend it would stabilize the enzyme, while still allowing the observed relaxed substrate specificity. Accordingly, we inserted 5 and 10 glycines in the

Downloaded from www.jbc.org at IReL (Trinity College Dublin), on April 2, 2011

Substrate Filtering in UCHL3



**FIGURE 5. The active site crossover loop restricts ubiquitin leaving group size.** *a*, cleavage of the crossover loop renders UCHL3 competent to disassemble ubiquitin polymers. UCHL3 Loop 2 protein (7 μg) was exposed to either 2.25 or 22.5 μg of SrtA and probe 1 (5 mM) for 2 h. Reactions were subsequently incubated with K63-linked diubiquitin chains (0.5 μg) at 37 °C for 1 h. Ubiquitin cleavage was assessed by 12.5% Tris/Tricine SDS-PAGE followed by anti-ubiquitin immunoblot. \*, Ub<sub>n</sub> denotes higher order K63-linked ubiquitin in the diubiquitin preparation. *b*, crossover loop expansion allows UCHL3-mediated hydrolysis of K63- and K48-linked ubiquitin but not a hexahistidine-tagged head-to-tail diubiquitin fusion. 180 ng of each enzyme (2, UCHL3 Loop 2; 5, 5 glycine insertion; 10, 10 glycine insertion; Iso-T, isopeptidase-T) was incubated with 1 μg of the indicated ubiquitin polymer for 2 h at 37 °C. Reactions were separated by 12.5% Tris/Tricine SDS-PAGE followed by anti-ubiquitin immunoblot. \*, Ub<sub>n</sub> denotes higher order K63-linked ubiquitin in the diubiquitin preparation. *c*, loop expanded versions of PfUCHL3 hydrolyze K48-linked diubiquitin, inefficiently cleave K63-linked polyubiquitin, and are inactive against a hexahistidine-tagged head-to-tail diubiquitin fusion. 180 ng of each enzyme (2, PfUCHL3 Loop 2; 5, 5 glycine insertion; 10, 10 glycine insertion; Iso-T, isopeptidase-T) was incubated with 1 μg of the indicated ubiquitin polymer for 2 h at 37 °C. Reactions were analyzed as in *b*.

are able to hydrolyze a linear hexahistidine-tagged head-to-tail diubiquitin fusion, whereas isopeptidase-T (USP5) efficiently does so. The PfUCHL3 Loop 2 mutant was also subjected to loop expansion and incubated with the various ubiquitin polymers. The PfUCHL3 loop-expanded mutants hydrolyze K48-, and to a lesser extent K63-linked ubiquitin, but not linear ubiquitin polymers (Fig. 5c). We assessed the rates of catalysis of our engineered mutant forms of UCHL3 and PfUCHL3 and compared these to USP8 (17) and the A20 N-terminal domain (18, 19) (Fig. 6). Quantitation of K48-linked diubiquitin hydrolysis yields rates intermediate between those of USP8 and the A20 N-terminal domain (Table 1). Thus, by expanding the length of the crossover loop, we have successfully created gain-of-function versions of UCHL3 and PfUCHL3 that can hydrolyze bulky substrates that the parent molecules are unable to attack.

We tested whether the loop-expanded versions of UCHL3 can hydrolyze the proximal ubiquitin from a ubiquitinated substrate, Mcl-1 (myeloid cell leukemia-1 protein). The HECT (homologous to E6-associated protein (E6AP) C terminus) domain E3-ligase, ARF-BP1/Mule, targets the antiapoptotic Bcl-2 (B-cell leukemia/lymphoma 2 protein) family member Mcl-1 for ubiquitination at several internal lysines *in vitro* (14). When such ubiquitinated Mcl-1 preparations<sup>3</sup> are exposed to the UCHL3 mutant with the 10 glycine expansion, the mono-ubiquitinated form of Mcl-1 is hydrolyzed (Fig. 7a). In contrast to wild-type UCHL3, substrates enjoy unfettered access to the active site of the M48<sup>USP</sup> isopeptidase. The M48<sup>USP</sup> domain, which has activity against K63- and K48- linked ubiquitin chains similar to that of the loop-expanded UCHL3 enzymes (13), also efficiently hydrolyzes the mono-ubiquitinated Mcl-1 substrate. To ensure that the observed multi-(mono)ubiquitinated forms of Mcl-1 indeed represent a single ubiquitin condensed onto several Mcl-1 lysines and not a ubiquitin chain added to a single lysine, we performed the Mcl-1 ubiquitination reaction with a ubiquitin variant incapable of polyubiquitin chain formation because it lacks lysine residues. Although the Mcl-1 ubiquitination pattern of the 13 available lysines in Mcl-1 is different with this ubiquitin variant, the UCHL3 mutant with 10 added glycines again hydrolyzes all proximal ubiquitin moieties, as assessed by a dramatic increase in the amount of Mcl-1 backbone released after digestion (Fig. 7b).

**DISCUSSION**

We took advantage of the known crystal structures for liganded and unliganded UCHL3 and applied the sortagging technique to a mammalian and apicomplexan representative of UCHL3. We chose to compare PfUCHL3 with its mammalian homolog because their sequences are no more than 35.9% identical, with almost no similarity in the crossover loop (20). Thus, the sequence-specific contributions of the crossover loop to catalysis can be parsed out of its structural contributions by comparison of both the host and pathogen enzymes. We have meanwhile solved the PfUCHL3 structure and found the overall fold to be virtually identical to its human counterpart.<sup>4</sup> Only the unliganded form of UCHL3 is a substrate for transpeptidation; UCHL3 was not cleaved when modified with HA-UbVME prior to exposure to sortase. Combined, our results suggest that the crossover loop of UCHL3 is indeed a flexible structure, despite the ordered loop visible in the crystal structure of its relative, UCHL1. Because there is no obvious sequence similarity between the crossover loops of UCHL3 and PfUCHL3, the presence, but not the exact sequence of the crossover loop is essential to substrate selection. This conclusion is supported also by the lack of effect of the introduction of an LPETG tag at

Downloaded from www.jbc.org at IReL (Trinity College Dublin) on April 2, 2011

Substrate Filtering in UCHL3

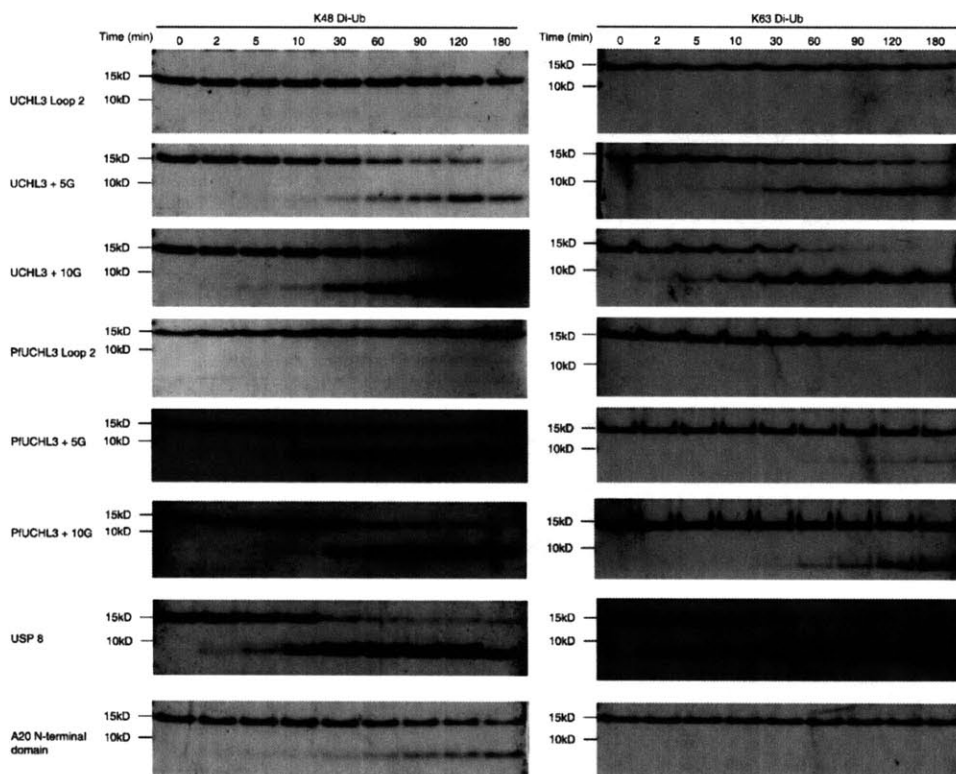


FIGURE 6. Isopeptide-linked diubiquitin hydrolysis by engineered UCHL3 and PfUCHL3 mutants. Enzymes (500 nM) were incubated with either K48-linked diubiquitin (10  $\mu$ M, left) or K63-linked diubiquitin (10  $\mu$ M, right), stopped at the indicated time points by addition of Tris/Tricine sample buffer, subjected to 10% Tris/Tricine SDS-PAGE electrophoresis, and stained by Colloidal Blue. K48-linked diubiquitin hydrolysis was quantified, and rates were determined (Table 1). Because of visible higher order ubiquitin polymers, K63-linked ubiquitin hydrolysis is displayed, but not quantitated.

**TABLE 1**  
K48-diubiquitin hydrolysis rates of engineered UCHL3 and PfUCHL3 enzymes compared to USP8 and the A20 N-terminal domain

To determine velocities, K48-diubiquitin cleavage reactions were separated by SDS-PAGE (Fig. 6), stained with colloidal Coomassie Blue, and the ratio of monoubiquitin to total ubiquitin was quantitated. The rate of monoubiquitin release versus time was determined by fitting the initial linear data points to a least squares regression line.

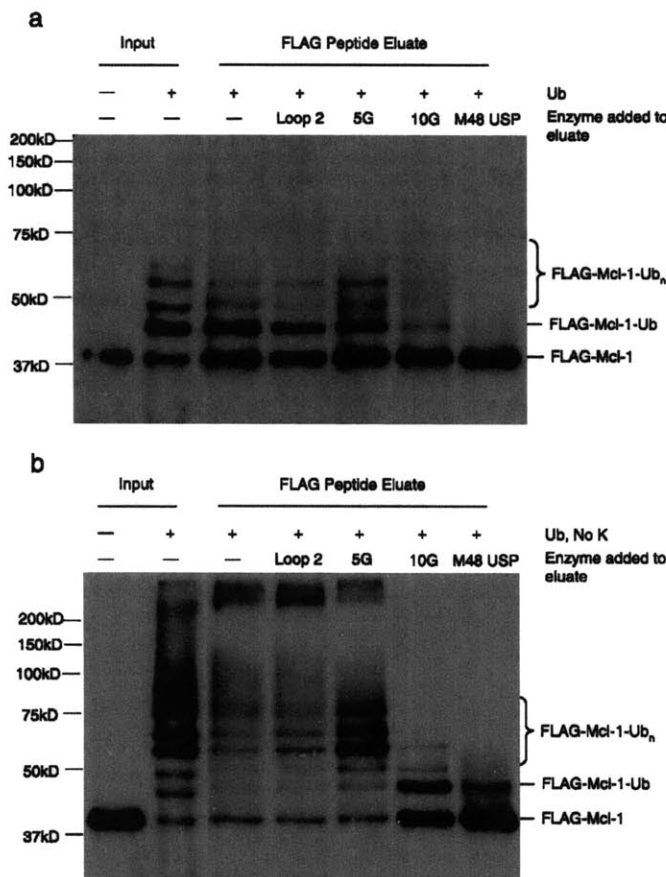
Enzyme	V/[E]	Rate relative to USP8	Rate relative to A20 NTD <sup>a</sup>
	min <sup>-1</sup>	%	%
UCHL3 + 5G	0.281	27.89	195.15
UCHL3 + 10G	0.410	40.61	284.19
PfUCHL3 + 5G	0.311	30.80	215.53
PfUCHL3 + 10G	0.374	37.05	259.22
USP8	1.009	100	700
A20 NTD	0.144	14.27	100

<sup>a</sup> NTD, N-terminal domain.

different positions in the crossover loop. The loop alone determines the size of the leaving group that can be liberated from ubiquitin by UCHL3. Integrity of the loop is not required for catalysis by UCHL3—the enzyme can function with its backbone cleaved very close to the active site. The stability of the

enzyme is likely enhanced by the complex and knotted topology of the polypeptide backbone. There are many examples of proteins that are cleaved in unstructured loops; such cleavage is often required to generate an active enzyme from its inactive precursor, to expose the fusogenic properties of viral fusion proteins (21), or to render active the bacterial toxins of the AB type such as cholera toxin and *E. coli* heat labile enterotoxin (22). The cleavage of unstructured loops does not, as a rule, perturb secondary or tertiary structure for those proteins where the two forms can be examined independently. However, the processed proteins but not their precursors are often poised to undergo major structural rearrangements upon engagement of the appropriate counterstructures.

We have obtained gain-of-function mutants of UCHL3 through enlargement of the active site crossover loop. These mutants successfully disassemble both K48 and K63  $\epsilon$ -amine linked ubiquitin polymers, while wild-type UCHL3, which has a crossover loop diameter of  $\sim$ 12–15 Å, cannot. When known structures for K48- and K63-linked ubiquitin dimers are modeled in the solved UCHL3-UbVME structure (supplemental



**FIGURE 7. Crossover loop expansion allows hydrolysis of ubiquitinated Mcl-1.** *a*, loop expanded UCHL3 hydrolyzes ubiquitin-FLAG-Mcl-1 conjugates. FLAG-tagged Mcl-1 (1  $\mu$ g) was ubiquitinated by incubation with UBE1, UBCH7, ArfBP-1/Mule, and ubiquitin (see "Experimental Procedures"). Mcl-1 was then immunoprecipitated with anti-FLAG antibody and either boiled in sample buffer (input) or eluted with FLAG peptide (FLAG peptide eluate). FLAG peptide eluates were pooled, divided equally, and incubated with 800 ng of the indicated enzymes for 2 h at 37 °C. Reactions were separated by 10% SDS-PAGE and analyzed by anti-Mcl-1 immunoblot. *b*, loop expanded UCHL3 hydrolyzes the proximal ubiquitin from Mcl-1. FLAG-tagged Mcl-1 was prepared as in *a*, using ubiquitin with all lysines substituted with arginine. Mcl-1 conjugates were digested and analyzed as in *a*.

Fig. S2), the proximal ubiquitin sterically clashes with UCHL3. To explain the observed hydrolysis of these substrates, the hinge region that contains the scissile Gly-Lys bond must be sufficiently flexible in solution to allow displacement of the proximal ubiquitin by UCHL3, consistent with previous observations (23). Because K63 lies near the N terminus of ubiquitin,  $\alpha$ -linked dimers are hypothesized to have an open conformation similar to that of K63-linked diubiquitin (24). However, none of the UCHL3 or pUCHL3 mutants can hydrolyze an  $\alpha$ -amine-linked dimer. This does not result from a difference between the  $\alpha$  and  $\epsilon$  linkage itself, because UCHL3 can hydrolyze a ubiquitin fusion to the ribosomal protein CEP52, in

### Substrate Filtering in UCHL3

$\alpha$ -linkage (4). Instead, the scissile bond in the ubiquitin head-to tail fusion is likely inaccessible to the UCHL3 active site cysteine. When the head-to tail ubiquitin dimer is modeled by ISG15, which has two tandem  $\alpha$ -linked Ubl domains, and fitted into the UCHL3-UbVME structure, the region where the scissile bond would be lies far outside of the active site groove (supplemental Fig. S2). Although the structural details of ISG15 likely differ from those of a ubiquitin head-to-tail fusion, we favor a model where the lack of cleavage by UCHL3 is due to differences in the conformation of K48-, K63-, and  $\alpha$ -amine-linked ubiquitin dimers in solution.

The promiscuity of the loop-expanded UCHL3 variants provides a rationale for the properties of UCHL3. In the absence of other protein domains that target the enzyme to particular ubiquitinated proteins or ubiquitin chains of specific linkages (25), UCHL3 limits access to its active site to only those substrates that can pass through the narrow bore of the crossover loop. Thus it is unlikely that wild-type UCHL3 directly targets larger, folded proteins for deubiquitination (26).

In summary, we have reported a novel method for exploring the contribution of flexible loops to protein structure and function. By applying the sortagging technique to the active site crossover loop in UCHL3 as well as mutagenesis, we can examine the contributions of the loop to both the structure and function of the enzyme. We have expanded the range of substrates that UCHL3 can hydrolyze and

provide strong support for the notion that the unusual active site crossover loop functions as a substrate filter, limiting the types of substrates that the enzyme can hydrolyze. Such modified versions of UCHL3 may be useful as general ubiquitin releasing enzymes for the study and identification of ubiquitin adducts.

*Acknowledgments*—We thank K. Love and R. Sastry for assistance with Mcl-1 ubiquitination reactions and purified FLAG-Mcl-1, C. Schlieker for purified M48<sup>USP</sup>, and G. Grotenbreg for the biotinylated nucleophile peptide.



*Chapter 5: Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations*

**Substrate Filtering in UCHL3**

**REFERENCES**

1. Kerscher, O., Felberbaum, R., and Hochstrasser, M. (2006) *Annu. Rev. Cell Dev. Biol.* **22**, 159–180
2. Love, K. R., Catic, A., Schlieker, C., and Ploegh, H. L. (2007) *Nat. Chem. Biol.* **3**, 697–705
3. Nijman, S. M., Luna-Vargas, M. P., Velds, A., Brummelkamp, T. R., Dirac, A. M., Sixma, T. K., and Bernards, R. (2005) *Cell* **123**, 773–786
4. Larsen, C. N., Krantz, B. A., and Wilkinson, K. D. (1998) *Biochemistry* **37**, 3358–3368
5. Johnston, S. C., Riddle, S. M., Cohen, R. E., and Hill, C. P. (1999) *EMBO J.* **18**, 3877–3887
6. Das, C., Hoang, Q. Q., Kreinbring, C. A., Luchansky, S. J., Meray, R. K., Ray, S. S., Lansbury, P. T., Ringe, D., and Petsko, G. A. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4675–4680
7. Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D., and Hill, C. P. (1997) *EMBO J.* **16**, 3787–3796
8. Misaghi, S., Galaray, P. J., Meester, W. J., Ovaa, H., Ploegh, H. L., and Gaudet, R. (2005) *J. Biol. Chem.* **280**, 1512–1520
9. Virnau, P., Mimy, L. A., and Kardar, M. (2006) *PLoS Comput. Biol.* **2**, e122
10. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E., and Ploegh, H. L. (2007) *Nat. Chem. Biol.* **3**, 707–708
11. Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Ploegh, H. L., and Kessler, B. M. (2002) *Chem. Biol.* **9**, 1149–1159
12. Ton-That, H., Liu, G., Maxmanian, S. K., Faull, K. F., and Schneewind, O. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12424–12429
13. Schlieker, C., Weihofen, W. A., Frijs, E., Kattenhorn, L. M., Gaudet, R., and Ploegh, H. L. (2007) *Mol. Cell* **25**, 677–687
14. Chen, D., Kon, N., Li, M., Zhang, W., Qin, J., and Gu, W. (2005) *Cell* **121**, 1071–1083
15. Ovaa, H., Kessler, B. M., Rolen, U., Galaray, P. J., Ploegh, H. L., and Malsucci, M. G. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2253–2258
16. Hassiepen, U., Eidhoff, U., Meder, G., Bulber, J. F., Hein, A., Bodendorf, U., Lorthiois, E., and Martoglio, B. (2007) *Anal. Biochem.* **371**, 201–207
17. Mizuno, E., Iura, T., Mukai, A., Yoshimori, T., Kitamura, N., and Komada, M. (2005) *Mol. Biol. Cell* **16**, 5163–5174
18. Lin, S. C., Chung, J. Y., Lamothe, B., Rajashankar, K., Lu, M., Lo, Y. C., Lam, A. Y., Darnay, B. G., and Wu, H. (2008) *J. Mol. Biol.* **376**, 526–540
19. Wertz, I. E., O'Rourke, K. M., Zhou, H., Eby, M., Aravind, L., Seshagiri, S., Wu, P., Wiesmann, C., Baker, R., Boone, D. L., Ma, A., Koonin, E. V., and Dixit, V. M. (2004) *Nature* **430**, 694–699
20. Frickel, E. M., Quesada, V., Muething, L., Gubbels, M. J., Spooner, E., Ploegh, H., and Artavanis-Tsakonas, K. (2007) *Cell Microbiol.* **9**, 1601–1610
21. Skehel, J. J., Cross, K., Steinhauer, D., and Wiley, D. C. (2001) *Biochem. Soc. Trans.* **29**, 623–626
22. Chinnapen, D. J., Chinnapen, H., Saslowsky, D., and Lencer, W. I. (2007) *FEMS Microbiol. Lett.* **266**, 129–137
23. Eddins, M. J., Varadan, R., Fushman, D., Pickart, C. M., and Wolberger, C. (2007) *J. Mol. Biol.* **367**, 204–211
24. Pickart, C. M., and Fushman, D. (2004) *Curr. Opin. Chem. Biol.* **8**, 610–616
25. Reyes-Turcu, F. E., Shanks, J. R., Komander, D., and Wilkinson, K. D. (2008) *J. Biol. Chem.* **283**, 19581–19592
26. Butterworth, M. B., Edinger, R. S., Ovaa, H., Burg, D., Johnson, J. P., and Frizzell, R. A. (2007) *J. Biol. Chem.* **282**, 37885–37893

Downloaded from www.jbc.org at IReL (Trinity College Dublin) on April 2, 2011



**Chapter 6: Sortase-catalyzed transformations  
that improve the properties of cytokines**

**Sortase-catalyzed transformations that improve the properties of cytokines**  
(from: Popp, M. W., Dougan, S. K., Chuang, T. Y., Spooner, E., and Ploegh, H. L. (2011)  
*Proc Natl Acad Sci U S A* **108**, 3169-3174)

### **Abstract**

Recombinant protein therapeutics often suffer from short circulating half-life and poor stability, necessitating multiple injections and resulting in limited shelf-life. Conjugation to polyethylene glycol chains (PEG) extends the circulatory half-life of many proteins, but the methods for attachment often lack specificity, resulting in loss of biological activity. Using four-helix bundle cytokines as an example, we present a general platform that uses sortase-mediated transpeptidation to facilitate site-specific attachment of PEG to extend cytokine half-life with full retention of biological activity. Covalently joining the N- and C- termini of proteins to obtain circular polypeptides, again executed using sortase, increases thermal stability. We combined both PEGylation and circularization by exploiting two distinct sortase enzymes and the use of a molecular suture that allows both site-specific PEGylation and covalent closure. The method developed is general, uses a set of easily accessible reagents, and should be applicable to a wide variety of proteins, provided that their termini are not involved in receptor binding or function.

### **Introduction**

Many clinically relevant cytokines share a four-helix bundle structure, typified by IFN $\alpha$ 2, GCSF3, EPO, IL-2, IL-4, IL-7, IL-9 and IL-15. Co-crystal structures of cytokines with receptor fragments and biochemical studies that map residues critical for interaction of a cytokine with its receptor show that the receptor contacts the sides of the helical bundles

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

<sup>1</sup>. This mode of interaction positions the N and C termini of the cytokine away from the receptor.

Polyethylene glycol chains attached to therapeutically important proteins increase circulatory half-life, reduce clearance by kidney filtration, reduce proteolysis, and reduce the generation of neutralizing antibodies <sup>2-4</sup>. The attachment of PEG commonly employs standard chemistries that target reactive amino acid side chains (e.g. cysteine, lysine). This strategy often generates a heterogeneous mixture in which multiple amino acids in the target are modified with a PEG chain, necessitating cumbersome separations and characterization <sup>5</sup>. Such PEGylated molecules often show decreased biological activity, likely due to attachment of a PEG chain to a residue important for interaction with the receptor <sup>6</sup>-this problem may be overcome by site-specific PEGylation. Although engineering of a carefully placed unpaired cysteine residue allows site-specific PEGylation <sup>7-8</sup>, this method must be tailored to the specific protein target.

Sortase A from *Staphylococcus aureus* (SrtA<sub>staph</sub>) is a thiol-containing transpeptidase that recognizes an LPXTG motif in multiple structurally unrelated substrates <sup>9</sup>. SrtA<sub>staph</sub> cleaves the peptide bond between the threonine and glycine residues with concomitant formation of a thioacyl enzyme intermediate that involves the catalytic cysteine and the substrate threonine. This acyl-enzyme is resolved by nucleophilic attack by the N terminus of an oligoglycine peptide, resulting in formation of an amide bond between the substrate protein and the incoming nucleophile <sup>10</sup>. Sortase A tolerates C-terminal extensions of the oligoglycine nucleophile, allowing diverse functionalized nucleophiles

to be installed site-specifically onto proteins equipped with an LPXTG motif<sup>11-12</sup>. The related *Streptococcus pyogenes* sortase accepts di-alanine based nucleophiles, which the *S. aureus* enzyme does not. This sortase (SrtA<sub>strep</sub>) cleaves the LPXTA motif between threonine and alanine and allows installation of modified alanine-based nucleophiles. SrtA<sub>strep</sub> also recognizes and cleaves LPXTG motifs, albeit with reduced efficiency, however the LPXTA motif is refractory to cleavage by SrtA<sub>staph</sub><sup>13-14</sup>.

Here we present a general strategy for site-specific modification of therapeutic recombinant proteins at their termini, an approach that is particularly well suited to the four-helix bundle cytokines, where the termini are distant from the receptor interaction site. We show improvements in thermal stability by covalently joining the N and C termini of these proteins<sup>15</sup>. We present a general method for combining both site-specific PEGylation and improved thermal stability in a single molecule by using a molecular suture produced in two successive rounds of sortase catalyzed transpeptidation (**Fig. 6.1**).

## Results

**Modification of IFN $\alpha$ 2.** We first applied the sortase reaction to human interferon alpha (IFN $\alpha$ 2), a four helix bundle cytokine where chemical PEGylation results in multiple positional isomers and a ~90% decrease in biological potency<sup>16</sup>. We expressed in *E. coli* two versions of IFN $\alpha$ 2, both with an LPETG sortase motif immediately followed by a hexahistidine tag for ease of purification. At the amino terminus, one of the two proteins possesses a diglycine motif for cyclization. This material can either be cyclized or

PEGylated at the C-terminus (**Figure 6.2a**). IFN $\alpha$ 2 with an N-terminal diglycine motif and a C-terminal LPETG was successfully cyclized by incubation with SrtA<sub>staph</sub> and purified to homogeneity by ion-exchange chromatography. The mass of the circularized product in a crude sortase reaction differs from a linear hydrolysis product by -18Da, consistent with an intramolecular transpeptidation reaction (**Figure 6.2b**). We identified the unique junction peptide that arises from intramolecular cyclization by MS/MS analysis of tryptic digests (**Figure 6.2c**). Site-specific PEGylation of the IFN $\alpha$ 2 variant that lacks the N-terminal diglycine motif (and thus unable to undergo intramolecular cyclization) was achieved by incubation with sortase and a GGGK peptide to which a 10 kDa PEG moiety is affixed via the lysine  $\epsilon$ -amine (Probe 1, **Figure 6.2a, Supplementary Figure 6.1a**). We confirmed equipotent biological activity of all four interferon alpha variants in a Daudi cell proliferation inhibition assay when compared to a commercially available, non-PEGylated preparation (**Figure 6.2d, Supplementary Table 1**). Addition of a protein-sized PEG module to the C-terminus of IFN $\alpha$ 2 does not appreciably perturb receptor binding and biological potency. We hypothesized that the circular IFN $\alpha$ 2 would be more resistant to thermal denaturation, because the termini are clamped shut and thus should not “fray”<sup>15, 17</sup>. We measured thermal denaturation with a ThermoFluor assay, using the commercially available Sypro Orange dye<sup>18</sup>. Indeed, the circular form of IFN $\alpha$ 2 had a significantly elevated  $T_m$  compared to the other three variants, all of which exhibited nearly identical  $T_m$ 's (**Figure 6.2e, Supplementary Table 6.1, Supplementary Figure 6.2a**). Finally, we injected mice with either the linear IFN $\alpha$ 2 variant used to construct the PEGylated form, or the IFN $\alpha$ 2 variant bearing a 10kDa PEG chain (Probe 1) at its C-terminus and measured the serum levels over time by ELISA. The PEGylated

form decayed with slower kinetics than its non-PEGylated counterpart (**Figure 6.2f**, **Supplementary Table 6.1**).

**Modification of a single protein preparation by multiple nucleophiles.** The sortase platform offers the ability to install nearly any non-genetically encoded entity onto the C-terminus of a single preparation of recombinant protein. We demonstrate this by synthesizing a nucleophile that bears a 20kDa PEG moiety (Probe 2, **Figure 6.1a**), as well as the previously described 10kDa PEG probe (Probe 1), and appended them to the identical preparation of IFN $\alpha$ 2 (**Figure 6.3a**). Both conjugates were tested for biological activity (**Figure 6.3b**) and are highly potent (**Supplementary Table 6.2**).

**Superior properties endowed upon sortase substrates are applicable to multiple proteins.** Sortase allows the installation of the same non-natural group onto different proteins, unrelated in sequence or amino acid composition. Accordingly, we applied the sortase reaction to a different four-helix bundle cytokine, GCSF-3, with similar results. We first made the linear precursor GCSF variants and then cyclized or PEGylated them with probe 1 using sortase (**Figure 6.4a**). All variants possess equal or superior *in vitro* bioactivity relative to a commercial non-PEGylated preparation, as assessed by a standard NFS-60 cell proliferation assay<sup>19</sup> (**Figure 6.4b**, **Supplementary Table 6.3**). Like cyclic IFN $\alpha$ 2, cyclic GCSF-3 shows increased thermostability (**Figure 6.4c**, **Supplementary Figure 6.2b**, **Supplementary Table 6.3**) and PEGylated GCSF-3 has a significantly extended circulatory half-life relative to its non-PEGylated precursor (**Figure 6.4d**, **Supplementary Table 6.3**) upon injection into mice. PEGylated GCSF-3 also led to a

more robust and prolonged proliferation of granulocytes in injected mice than its non-PEGylated analog (**Figure 6.4e**).

**Modification of glycoproteins.** Many important therapeutics are glycosylated proteins that traverse the secretory pathway and are produced in mammalian cell culture. Are such proteins equally amenable to engineering using a sortase-mediated transpeptidation reaction? We chose human erythropoietin (EPO) because ~40% of its mass is composed of bulky, charged N-linked glycans, and cyclization provides a stringent test of the sortase method, as it requires (minimal) modifications at both termini. We replaced the endogenous EPO signal sequence with the murine H-2K<sup>b</sup> signal sequence, followed by two glycine residues. This signal peptide is cleaved in mammalian cells with concomitant exposure of an N-terminal glycine residue. We added the requisite sortase motif followed by a hexahistidine tag at the C-terminus of EPO for ease of purification. HEK-293T cell lines stably transduced with this construct yielded preparations that were <50% pure after a single round of Ni-NTA IMAC of conditioned medium. Material from such preparations can be cyclized successfully upon incubation with sortase (**Figure 6.5a**). Again, we identified the unique junction peptide that arises from covalent ligation of the N and C termini by MS/MS analysis of of cyclized EPO, shorn of all glycans by incubation with PNGaseF, and digested with trypsin (**Figure 6.5b**). Both the linear preparations as well as the sortase reaction containing cyclic EPO were tested in a cell proliferation assay with Ba/F3 cells that stably express the erythropoietin receptor<sup>20</sup> and these compared favorably with commercial EPO preparations. Most importantly, the cyclic form is as potent as its linear counterpart (**Figure 6.5c, Supplementary Table**

6.4), indicating that modification of the EPO termini does not affect its biological activity.

**Two-step transacylation allows combination of enhanced properties.** Having shown for multiple examples that sortase-mediated site-specific modification of the C-terminus with PEG increases circulatory half-life with nearly no loss of biological activity, and that covalent closure of the N and C termini yield proteins stabilized against thermal denaturation, we executed a scheme that combines these desirable properties (**Figure 6.1**). In the first step, we exploit a previously described sortase from *Streptococcus pyogenes*<sup>13-14</sup> that accepts alanine-based nucleophiles to affix a peptide containing a non-natural amino-oxy containing amino acid, followed by the SrtA<sub>Staph</sub> cleavage site. We use SrtA<sub>Staph</sub> to effectuate covalent closure, as the LPETAA site left by SrtA<sub>Strep</sub> is resistant to attack by SrtA<sub>Staph</sub>, to yield a circular protein with the amino-oxy group at the place of suture. Next, the amino-oxy group is used in a bio-orthogonal oxime ligation reaction<sup>21-23</sup> with commercially available methoxy-capped PEG-propionaldehyde (**Supplementary Figure 6.1b**). We recombinantly expressed IFN $\alpha$ 2 bearing two glycines at the N-terminus, and at the C-terminus the SrtA<sub>Strep</sub> cleavage site, followed by a hexahistidine tag. The purified protein was subjected to the two-step transacylation procedure to yield circular IFN $\alpha$ 2 bearing the amino-oxy group. After ion-exchange chromatography, this material was PEGylated and purified by cation exchange chromatography (**Figure 6.6a**). Without optimization, the final yield of circular PEGylated protein was ~10%, because of the multiple separate chromatography and protein concentration steps involved in production and PEGylation. The mass of the



material before oxime ligation was consistent with its cyclization (**Supplementary Figure 6.3a**). Digestion with AspN and MS/MS analysis revealed the unique peptide that arises as a consequence of backbone cyclization consisting of the molecular suture probe stitched between the C- and N-termini of IFN $\alpha$  (**Supplementary Figure 6.6b**). All three variants yielded virtually indistinguishable IC<sub>50</sub> values in an inhibition of cell proliferation assay (**Figure 6.6c, Supplementary Table 6.5**) and the circular versions were stabilized against thermal denaturation (**Figure 6.6d, Supplementary Table 6.5, Supplementary Figure 6.2c**). The circular versions remained biologically potent after boiling when given the opportunity to refold, whereas we were unable to extract an IC<sub>50</sub> value for the boiled linear preparation over the concentration range assayed (**Figure 6.6e, Supplementary Table 6.6**). We injected the linear and the circular, PEGylated forms of IFN $\alpha$ 2 into the tail vein of mice and measured half life by ELISA. The circular, PEGylated species was cleared significantly more slowly than the linear form (**Figure 6.6f, Supplementary Table 6.5**).

## **Discussion**

The sortase transpeptidation reaction allows facile site-specific PEGylation of multiple distinct proteins. In all cases tested, the site-specific C-terminal PEGylation proceeds efficiently (**Supplementary Figure 6.3b**) and yields adducts of known stoichiometry that are biologically equipotent to the non-PEGylated versions, but retain the increase in circulatory half life associated with PEG modification. This we attribute to the fact that all enzymatic transformations are performed on native proteins, procedures that should not affect overall conformation or exposure of functionally important sidechains.

Because our approach to PEGylation is site-specific, all preparations are also homogenous and easily purified<sup>5</sup>. Although we limit our examples to four-helix bundle cytokines, this platform should be readily extended to structurally distinct therapeutic proteins, with the singular requirement that the C-terminus is not involved in receptor binding. A sortase-based approach requires the genetic fusion of a very small (5 amino acid) tag to the protein of interest and all transformations occur under native conditions. The additional amino acids that result from fusion to the sortase recognition motif are close to the site of PEG attachment. Immunogenicity of this site is very likely reduced by the PEG moiety shielding this area<sup>3</sup>. We and others have labeled proteins with an exposed N-terminal glycine using sortase<sup>14,24</sup>. This approach should therefore be readily extended to site-specific PEGylation at the N-terminus as well.

Cyclization of the backbone of protein-based therapeutics also proceeds efficiently (**Supplementary Figure 6.3b**) and yields preparations that are more resistant to thermal denaturation. Cyclic proteins are also resistant to exoproteolytic attack<sup>25-26</sup>, a feature that may enhance utility of any therapeutic proteins exposed to exoproteases, for example upon receptor mediated internalization. In addition, cyclization of proteins and peptides has been shown to improve potency, stability and oral bioavailability<sup>27-30</sup>.

Finally, we have inserted a non-template encoded entity in what is topologically internal to a circular protein, a feat that cannot be accomplished genetically by intein-based methods or by any other currently known means. More generally, this dual transacylation scheme can be used to insert non-natural groups between fully native

proteins expressed separately, or between protein domains. We suspect that this approach will find application in the protein engineering field and so further extend the utility of sortases as tools.

## **Methods**

**Sortase Reactions.** Sortase reactions with SrtA<sub>Staph</sub> were performed as described previously<sup>11</sup>. For either cyclization or PEGylation, reactions containing 50  $\mu$ M substrate, 50  $\mu$ M SrtA<sub>Staph</sub> and 1 mM probe (for PEGylation) were incubated overnight at 25°C without agitation. Reactions were purified by cation exchange chromatography on a Mono-S (GE Healthcare). For IFN $\alpha$ 2 reactions, continuous gradient chromatography was performed with 20mM MES pH 5.0 and 20mM MES, 1M NaCl pH5.0, as eluent. For GCSF reactions, continuous gradient chromatography was performed with 20mM sodium acetate pH 4.5 and 20mM sodium acetate, 1M NaCl pH 4.5, as eluent. Peaks containing the desired product were pooled, concentrated with Vivaspin 500 centrifugal concentrators (Sigma) and protein concentration was determined by Bradford Assay (Bio-Rad). For EPO cyclization, 300  $\mu$ l of Ni-NTA elutate (0.101 mg/mL total protein) was incubated with 100  $\mu$ M SrtA and sortase buffer in 400  $\mu$ l total volume at 25°C for 16 hours. For two-step transacylation reactions, 150  $\mu$ M substrate was incubated with 50  $\mu$ M SrtA<sub>Strep</sub> and 2 mM probe in 100 mM Tris, 150 mM NaCl pH 8.0 for 4 hours at 37°C. SrtA<sub>Strep</sub> activity was halted by incubation with 500  $\mu$ M E-64 on ice for 1 hour. Crude reaction mixtures were then supplemented with 10 mM imidazole and subjected to Ni-NTA chromatography (Qiagen) to remove prematurely cyclized material. Protein was eluted with 50 mM Tris, 150 mM NaCl, 500 mM imidazole and buffer exchanged into

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

SrtA<sub>Staph</sub> buffer (50mM Tris, 150mM NaCl, 10 mM CaCl<sub>2</sub>) with a PD10 desalting column (GE Healthcare). For cyclization, this material was incubated overnight with 50  $\mu$ M SrtA<sub>Staph</sub> and purified by cation exchange as described for other IFN $\alpha$ 2a reactions.

Concentrated, purified circular IFN $\alpha$ 2 bearing the AOAA group was then diluted with 1 volume of 50mM sodium acetate pH 4.5, 150 mM NaCl containing 2 mM methoxy-capped PEG-propionaldehyde (NANOCS) and 100 mM aniline (JT Baker) and incubated at 30°C for 3 hours without agitation. This reaction mixture was again purified by cation exchange chromatography, concentrated, and protein concentration was determined by Bradford assay.

**Thermal Denaturation Assays.** Thermal shift assays were performed in a Roche 480 lightcycler and fluorescence was measured with Ex 533nm, Em=610nm. For GCSF-3, 1  $\mu$ g of each variant was mixed with 1  $\mu$ l of 100x Sypro orange in 25  $\mu$ l total volume of 50mM acetate pH 4.0, 150mM NaCl. n=4 for all samples with buffer as a blank control and lysozyme as a positive control. For IFN $\alpha$ 2, 1  $\mu$ g of each variant was mixed with 1  $\mu$ l of 100x Sypro orange (Invitrogen) in 25  $\mu$ l total volume of 50mM MES ph 5.0, 150mM NaCl. n=4 for all samples with buffer as a blank control and lysozyme as a positive control. Samples were heated from ambient temperature to 95°C at 0.01°C/s with 6 acquisitions/°C. The included Roche software was used to calculate the negative first derivative of the temperature change as a function of time and the minima were used as the T<sub>m</sub>. Because of the impurities in the cyclic EPO preparations, T<sub>m</sub> was not determined.

**Cell Based Bioactivity Assays.** Daudi cell proliferation inhibition assays to measure the activity of IFN $\alpha$ 2 conjugates were performed as described<sup>8</sup> and compared to commercial IFN $\alpha$ 2 (PBL Interferon Source) with low passage Daudi cells (ATCC), cultured in RPMI/10%IFS/50 units/ml penicillin, 50  $\mu$ g/ml streptomycin sulfate. Cell proliferation was measured by MTT assay according to manufacturer's directions (ATCC). For the denaturation assay, samples diluted to 1  $\mu$ g/ml in RPMI/10% IFS were boiled for 4 min, allowed to cool at room temperature for 16 hours, and assayed for Daudi cell proliferation inhibition. NFS-60 cell proliferation assays to measure G-CSF conjugate activity were performed as follows. NFS-60 cells (a kind gift from Dr. James Ihle, St. Jude Children's Research Hospital, Tennessee) were cultured in RPMI /10%IFS/50 units/ml penicillin, 50  $\mu$ g/ml streptomycin sulfate supplemented with murine IL-3 (20U/mL, R&D Systems). Cells were washed extensively in complete RPMI medium lacking IL-3 and resuspended at  $1 \times 10^5$  cells/ml in complete RPMI lacking IL-3. Titrated GCSF conjugates (50  $\mu$ l) in complete RPMI were aliquoted into a flat bottom 96 well plate and 50  $\mu$ l of cells ( $0.5 \times 10^4$  cells/well) were added to each well. Cells were cultured for 3 days and an MTT assay (ATCC) was performed according to manufacturer's directions. Each plate contained a titration of commercial GCSF3 (Peprotech) and readings were blanked against wells containing only NFS-60 cells. The activity of each conjugate was measured a minimum of n=3 times. BaF3 cells stably expressing the EPO receptor (BaF3 EPO-R cells, a kind gift from Dr. Harvey Lodish, Whitehead Institute) were used to measure EPO activity. BaF3 EPO-R cells were cultured as described for NFS-60 cells and assays were performed essentially as described for G-CSF, except EPO circularization reactions incubated with or without sortase for 16 hours at 37°C were

used. The total protein concentration of the EPO input material (measured by the Bradford method) was used and was not adjusted for impurities. Commercial preparations of EPO (eBioscience) were used as standards and readings were blanked against BaF3 EPO-R cells cultured with no cytokines. Activities were measured a minimum of n=3 times.

**Circulatory Half-life Assays.** For circulatory half-life assays, mice were injected in the tail vein with each protein (10  $\mu$ g per mouse for IFN $\alpha$ 2 conjugates, 5  $\mu$ g per mouse for G-CSF conjugates) and subjected to retro-orbital eye bleed at the indicated time points. Blood was harvested, centrifuged at 5000 rpm in a tabletop centrifuge, and serum was collected and snap-frozen. Elisa assays to measure the quantity of cytokine in serum samples were performed according to the manufacturer's directions (IFN $\alpha$  from PBL Interferon Source, G-CSF from Invitrogen). Each conjugate was injected into n=3 mice. Data was fit to a two phase exponential decay in GraphPad Prism.

**Granulocyte Proliferation Assay.** Peripheral blood was collected retroorbitally into EDTA collection tubes. Red blood cells were lysed in hypotonic lysis buffer and the remaining peripheral blood mononuclear cells were stained with anti-Gr-1-PE (BD Pharmingen), anti-CD11b-FITC (BD Pharmingen) and 7-AAD (ViaProbe, BD Pharmingen). Cells were analyzed using a FACS Caliber flow cytometer (BD).

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

**Acknowledgements.** We thank Drs. Nick Yoder, John Antos, and Annemmarthe van der Veen, for helpful discussions, and Drs. Harvey Lodish and James Ihle for reagents. This work was supported by grants from the N.I.H.

**Author Contributions.** M.W.P. designed, performed and analyzed research and wrote the paper. S.K.D. and E.S. performed and analyzed research. T.Y.C. performed research. H.L.P. designed research and wrote the paper.

## References

1. Wang, X., Lupardus, P., Laporte, S.L. & Garcia, K.C. Structural biology of shared cytokine receptors. *Annu Rev Immunol* **27**, 29-60 (2009).
2. Leader, B., Baca, Q.J. & Golan, D.E. Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* **7**, 21-39 (2008).
3. Hershfield, M.S. et al. Use of site-directed mutagenesis to enhance the epitope-shielding effect of covalent modification of proteins with polyethylene glycol. *Proc Natl Acad Sci U S A* **88**, 7185-7189 (1991).
4. Jevsevar, S., Kunstelj, M. & Porekar, V.G. PEGylation of therapeutic proteins. *Biotechnol J* **5**, 113-128 (2010).
5. Veronese, F.M. Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials* **22**, 405-417 (2001).
6. Grace, M.J. et al. Site of pegylation and polyethylene glycol molecule size attenuate interferon-alpha antiviral and antiproliferative activities through the JAK/STAT signaling pathway. *J Biol Chem* **280**, 6327-6336 (2005).
7. Doherty, D.H. et al. Site-specific PEGylation of engineered cysteine analogues of recombinant human granulocyte-macrophage colony-stimulating factor. *Bioconjug Chem* **16**, 1291-1298 (2005).
8. Rosendahl, M.S. et al. A long-acting, highly potent interferon alpha-2 conjugate created using site-specific PEGylation. *Bioconjug Chem* **16**, 200-207 (2005).
9. Marraffini, L.A., Dedent, A.C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* **70**, 192-221 (2006).
10. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci U S A* **96**, 12424-12429 (1999).
11. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**, 707-708 (2007).
12. Tanaka, T., Yamamoto, T., Tsukiji, S. & Nagamune, T. Site-specific protein modification on living cells catalyzed by Sortase. *Chembiochem* **9**, 802-807 (2008).
13. Race, P.R. et al. Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J Biol Chem* **284**, 6924-6933 (2009).
14. Antos, J.M. et al. Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* **131**, 10800-10801 (2009).
15. Antos, J.M. et al. A straight path to circular proteins. *J Biol Chem* **284**, 16028-16036 (2009).
16. Bailon, P. et al. Rational design of a potent, long-lasting form of interferon: a 40 kDa branched polyethylene glycol-conjugated interferon alpha-2a for the treatment of hepatitis C. *Bioconjug Chem* **12**, 195-202 (2001).
17. Camarero, J.A. et al. Rescuing a destabilized protein fold through backbone cyclization. *J Mol Biol* **308**, 1045-1062 (2001).



18. Lo, M.C. et al. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **332**, 153-159 (2004).
19. Hara, K. et al. Bipotential murine hemopoietic cell line (NFS-60) that is responsive to IL-3, GM-CSF, G-CSF, and erythropoietin. *Exp Hematol* **16**, 256-261 (1988).
20. Zhang, Y.L. et al. Symmetric signaling by an asymmetric 1 erythropoietin: 2 erythropoietin receptor complex. *Mol Cell* **33**, 266-274 (2009).
21. Dirksen, A., Hackeng, T.M. & Dawson, P.E. Nucleophilic catalysis of oxime ligation. *Angew Chem Int Ed Engl* **45**, 7581-7584 (2006).
22. Kochendoerfer, G.G. et al. Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* **299**, 884-887 (2003).
23. Shao, H. et al. Site-specific polymer attachment to a CCL-5 (RANTES) analogue by oxime exchange. *J Am Chem Soc* **127**, 1350-1351 (2005).
24. Yamamoto, T. & Nagamune, T. 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 (2009).
25. Andersen, A.S. et al. Backbone cyclic insulin. *J Pept Sci* **16**, 473-479 (2010).
26. Iwai, H. & Pluckthun, A. Circular beta-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett* **459**, 166-172 (1999).
27. Clark, R.J. et al. Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. *Proc Natl Acad Sci U S A* **102**, 13767-13772 (2005).
28. Clark, R.J. et al. The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew Chem Int Ed Engl* **49**, 6545-6548 (2010).
29. Trabi, M. & Craik, D.J. Circular proteins--no end in sight. *Trends Biochem Sci* **27**, 132-138 (2002).
30. Craik, D.J. Chemistry. Seamless proteins tie up their loose ends. *Science* **311**, 1563-1564 (2006).
31. Trotter, M.D., Newsted, M.M., King, L.E. & Fraker, P.J. Natural glucocorticoids induce expansion of all developmental stages of murine bone marrow granulocytes without inhibiting function. *Proc Natl Acad Sci U S A* **105**, 2028-2033 (2008).
32. Lilley, B.N. & Ploegh, H.L. A membrane protein required for dislocation of misfolded proteins from the ER. *Nature* **429**, 834-840 (2004).

## Figure Legends

### Figure 6.1. The sortase reaction scheme.

Cyclization of substrates equipped with a C-terminal LPXTG sortase recognition element as well as an N-terminal glycine by SrtA<sub>Staph</sub> (left). Substrates lacking the N-terminal glycine and incubated in the presence of exogenous oligo-glycine functionalized PEG are site-specifically PEGylated (right). Circular, PEGylated proteins can be constructed by incubating substrates bearing an N-terminal glycine and a C-terminal SrtA<sub>Strep</sub> cleavage site (LPXTA) with an alanine-based nucleophile carrying an aminoxy group and the SrtA<sub>Staph</sub> cleavage site (Middle). This transpeptidation product is cyclized upon incubation with SrtA<sub>Staph</sub>. Finally, PEG is attached by aniline-catalyzed oxime ligation with methoxy-capped PEG propionaldehyde. The oxime bond formed is shown (inset).

### Figure 6.2. Synthesis and characterization of human IFN $\alpha$ 2 conjugates.

- (a) GG-IFN $\alpha$ -LPETGGHis<sub>6</sub> was used to obtain backbone-cyclized IFN $\alpha$ 2, and IFN $\alpha$ -LEPTGGHis<sub>6</sub> was used to make IFN $\alpha$ -PEG(10kDa), bearing a C-terminal PEG chain. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie.
- (b) ESI-MS characterization of IFN $\alpha$  species in crude sortase reactions either before (linear GG-IFN $\alpha$ -LPETGGHis<sub>6</sub>) or after (cyclic IFN $\alpha$ 2) overnight incubation at 25°C.
- (c) MS/MS identification of a tryptic peptide comprising the IFN $\alpha$ 2 C-terminus, followed by the SrtA<sub>Staph</sub> cleavage site, joined to the N-terminal glycines in the IFN $\alpha$ 2 precursor.
- (d) *In vitro* bioactivity of IFN $\alpha$  conjugates in a Daudi cell proliferation inhibition assay. Dose-response curves for Daudi cell proliferation inhibition were measured by (3-(4,5-

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium) bromide (MTT) assay, in triplicate, with the standard deviation displayed.

(e) Resistance of backbone-cyclized IFN $\alpha$  to thermal denaturation. Change in  $T_m$  for circular and PEGylated IFN $\alpha$  conjugates, relative to their precursor proteins, were measured by thermal shift assay, with n=4.

(f) Change in serum IFN $\alpha$  levels following tail vein injection of site-specifically PEGylated IFN $\alpha$  as well as its precursor protein. 10  $\mu$ g of each protein was injected and, at the indicated time points, blood was withdrawn and cytokine concentration in serum was determined by ELISA assay. Data are means  $\pm$  standard deviation for three mice in each group.

**Figure 6.3. Sortase installs multiple probes onto a single preparation of protein.**

(a) The same preparation of IFN $\alpha$ -LEPTGGHis<sub>6</sub> was PEGylated by sortase reaction with either a 10kDa or a 20kDa PEG nucleophile. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie.

(b) *In vitro* bioactivity of 10 kDa and 20 kDa PEGylated IFN $\alpha$  conjugates in a Daudi cell proliferation inhibition assay. Dose-response curves for Daudi cell proliferation inhibition were measured by MTT assay, in triplicate, with the standard deviation displayed.

**Fig. 6.4. Synthesis and characterization of human GCSF-3 conjugates.**

- (a) GG-GCSF-LPETGGHis<sub>6</sub> was used to make backbone cyclized GCSF-3, and GCSF-LEPTGGHis<sub>6</sub> was used to make GCSF-PEG(10kDa), bearing a C-terminal PEG chain. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie.
- (b) *In vitro* bioactivity of GCSF-3 conjugates in an NFS-60 cell proliferation assay. Dose-response curves for NFS-60 cell proliferation were measured by MTT assay, in triplicate, with the standard deviation displayed.
- (c) Resistance of backbone cyclized GCSF-3 to thermal denaturation. Change in  $T_m$  for circular and PEGylated GCSF-3 conjugates, relative to their precursor proteins, were measured by thermal shift assay (n=4).
- (d) Change in serum GCSF levels following tail vein injection of site-specifically PEGylated GCSF-3, as well as its precursor. 5 µg of each protein was injected and, at the indicated time points, blood was withdrawn and cytokine concentration in serum was determined by ELISA assay. Data are means from three mice in each group, with the standard deviation displayed.
- (e) GCSF, PEGylated via sortase, is highly potent *in vivo*. C57BL/6 mice (n=4 per group) were injected intravenously with linear GCSF, GCSF-PEG(10 kD), or saline control (PBS). Peripheral blood was collected at the indicated times and stained with antibodies to Gr-1 and CD11b. Newly generated granulocytes were defined as Gr-1 intermediate CD11b<sup>+</sup> cells<sup>31</sup>.

**Fig. 6.5. Synthesis and characterization of cyclic human erythropoietin.**

(a) GG-EPO-LPETGGHis<sub>6</sub> bearing the H-2Kb signal sequence was expressed in HEK-293T cells. Conditioned media was subjected to Ni-NTA IMAC purification. Eluted material was supplemented with 100  $\mu$ M SrtA<sub>Staph</sub> and incubated at 25°C. Aliquots were removed at the indicated time points, subjected to 12.5% SDS-PAGE and protein was visualized by silver staining.

(b) MS/MS identification of a peptide generated by tryptic digestion containing the C-terminus of EPO, followed by the SrtA<sub>Staph</sub> cleavage site and joined to the N-terminus of EPO.

(c) *In vitro* bioactivity of linear and cyclic EPO conjugates in BaF3-EPOR cell proliferation assay. Eluted material from Ni-NTA IMAC purification was incubated in the presence or absence of sortase for 16 h at 25°C and these crude reactions were measured for *in vitro* EPO bioactivity. Dose response curves for BaF3-EPOR cell proliferation were measured by MTT assay, in triplicate, with the standard deviation displayed. Concentrations were not adjusted for impurities in the EPO preparations.

**Fig. 6.6. Synthesis and characterization of cyclic, PEGylated IFN $\alpha$ 2.**

(a) GG-IFN $\alpha$ -LPETAHis<sub>6</sub> was used to make backbone-cyclized IFN $\alpha$ 2 with the aminoxy moiety stitched between the site of closure. This material was then used to generate cyclic, PEGylated IFN $\alpha$ 2 by aniline catalyzed oxime ligation. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie.

(b) MS/MS identification of a peptide generated by AspN digestion containing the C-terminus of IFN $\alpha$ , followed by the alanine-based probe joined to the N-terminus of IFN $\alpha$ .

by SrtA<sub>Staph</sub>. Note that the aminoxy group was likely lost during MS/MS analysis and ESI-MS reconstructions of unfragmented protein revealed a mass consistent with installation of the aminoxy group (**Supplementary Figure 6.3**).

- (c) *In vitro* bioactivity of linear and cyclic IFN $\alpha$  conjugates in a Daudi cell proliferation inhibition assay. Dose-response curves for Daudi cell proliferation inhibition were measured by MTT assay, in triplicate, with the standard deviation displayed.
- (d) Resistance of circular-AOAA and circular-PEGylated IFN $\alpha$  to thermal denaturation. Change in  $T_m$ 's for circular-AOAA and circular-PEGylated IFN $\alpha$  conjugates, relative to their precursor protein, were measured by thermal shift assay (n=4).
- (e) Circular cytokines remain biologically active after boiling. The indicated variants were diluted to 1mg/ml in RPMI/10% serum, boiled for 4 minutes, allowed to cool at room temperature for 16 hours, and assayed for inhibition of Daudi cell proliferation.
- (f) Change in serum IFN $\alpha$  levels following tail vein injection of circular, site-specifically PEGylated IFN $\alpha$  as well as its linear precursor protein. 10 $\mu$ g of each protein was injected and, at the indicated time points, blood was withdrawn and cytokine concentration in serum was determined by ELISA assay. Data are means from three mice in each group, with the standard deviation displayed.

Figure 6.1

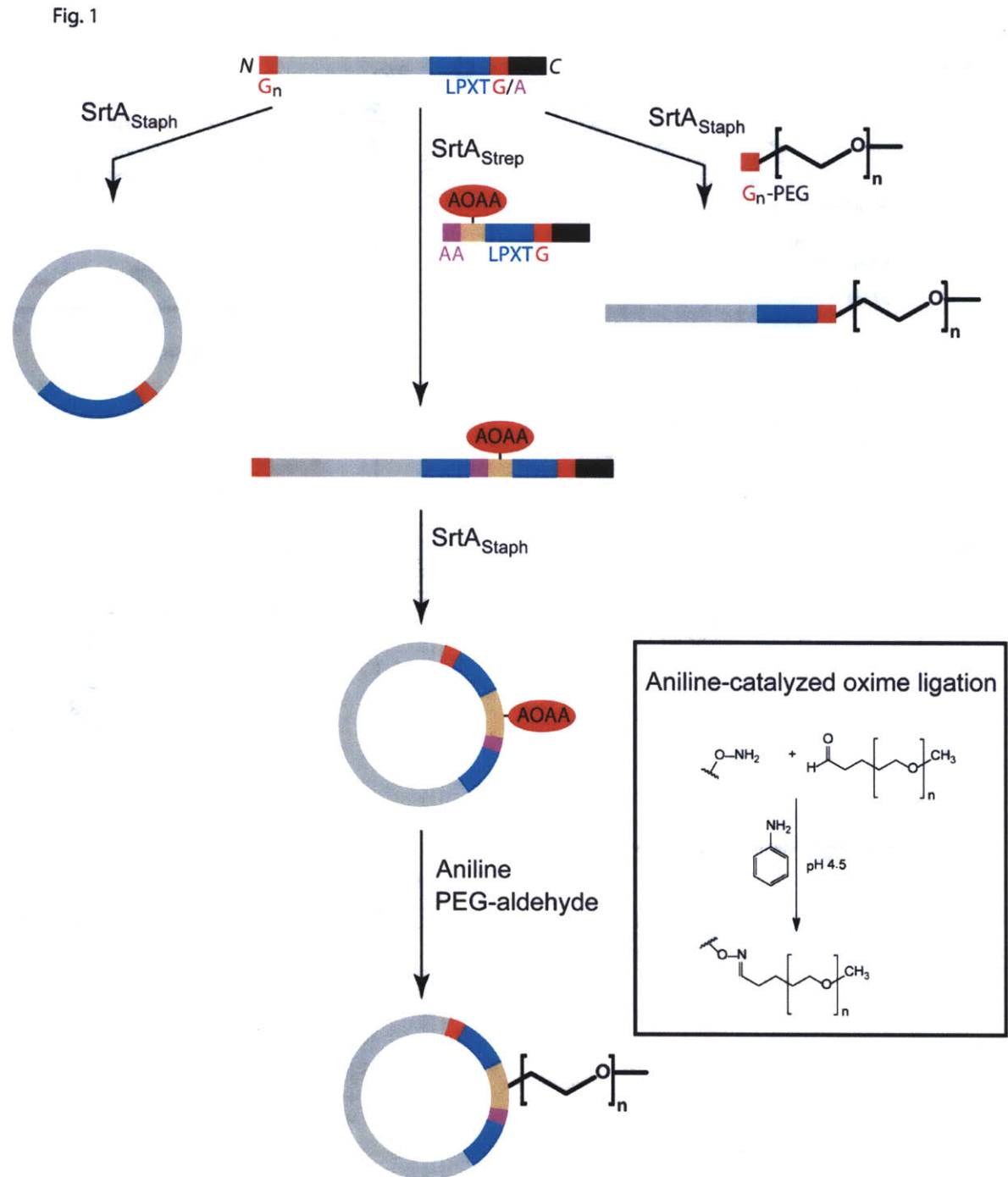


Figure 6.2

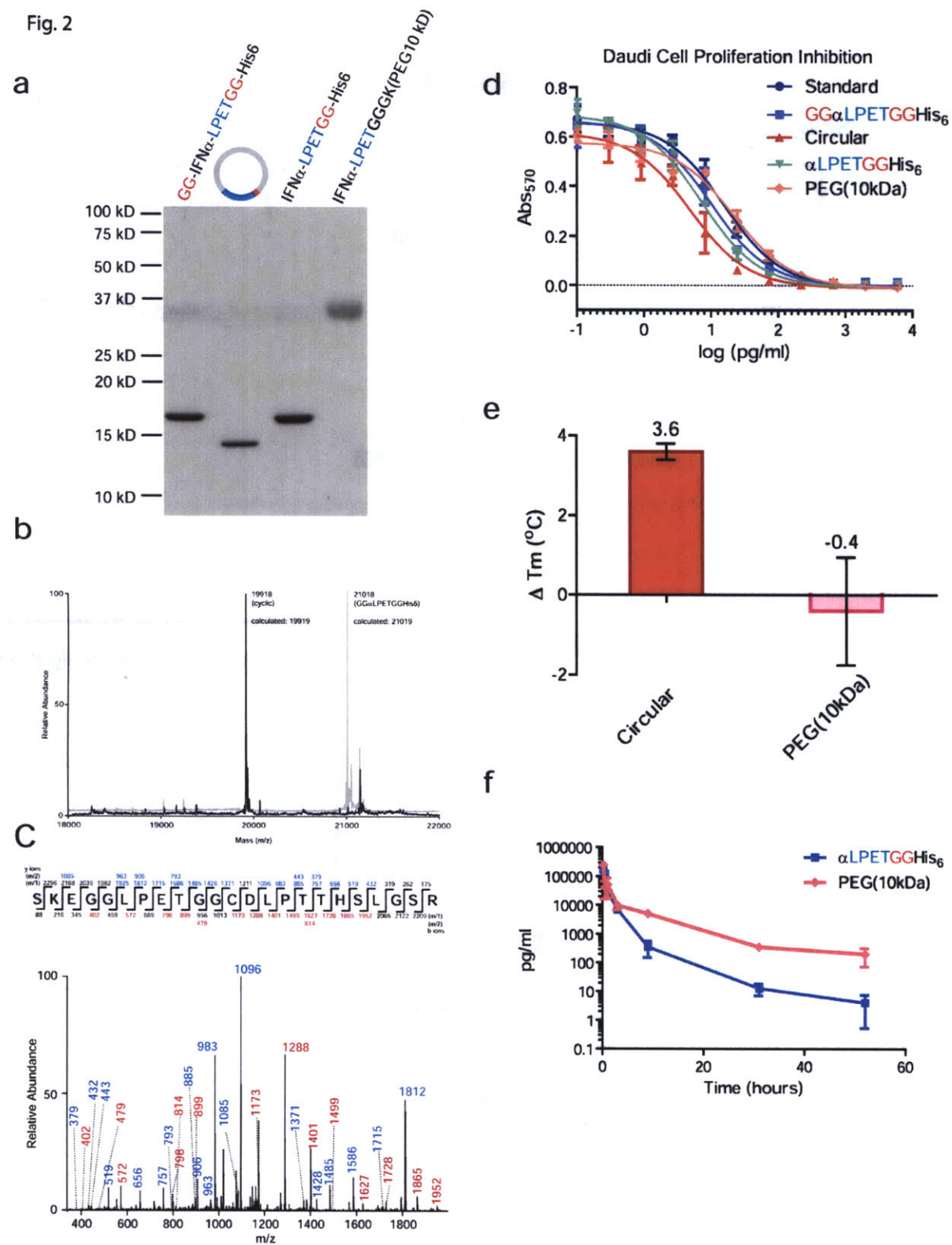
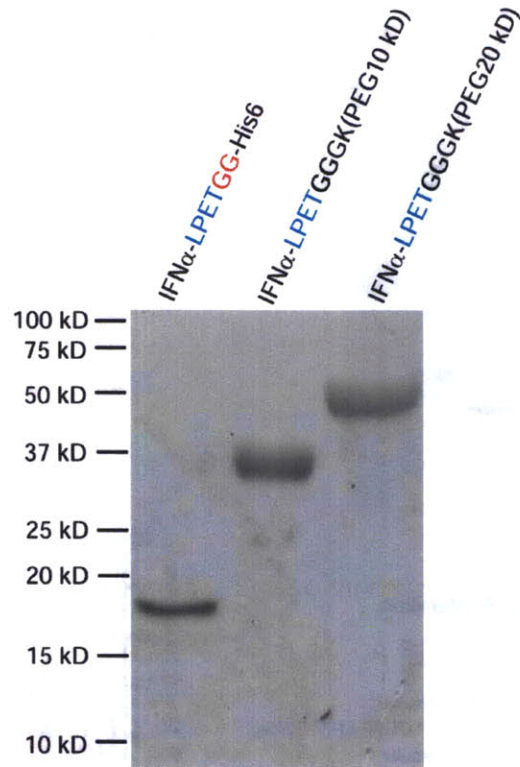




Figure 6.3

Fig. 3

a



b

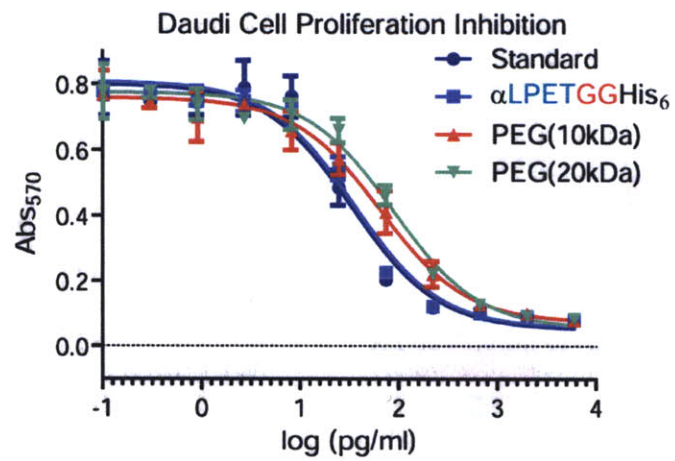


Figure 6.4

Fig. 4

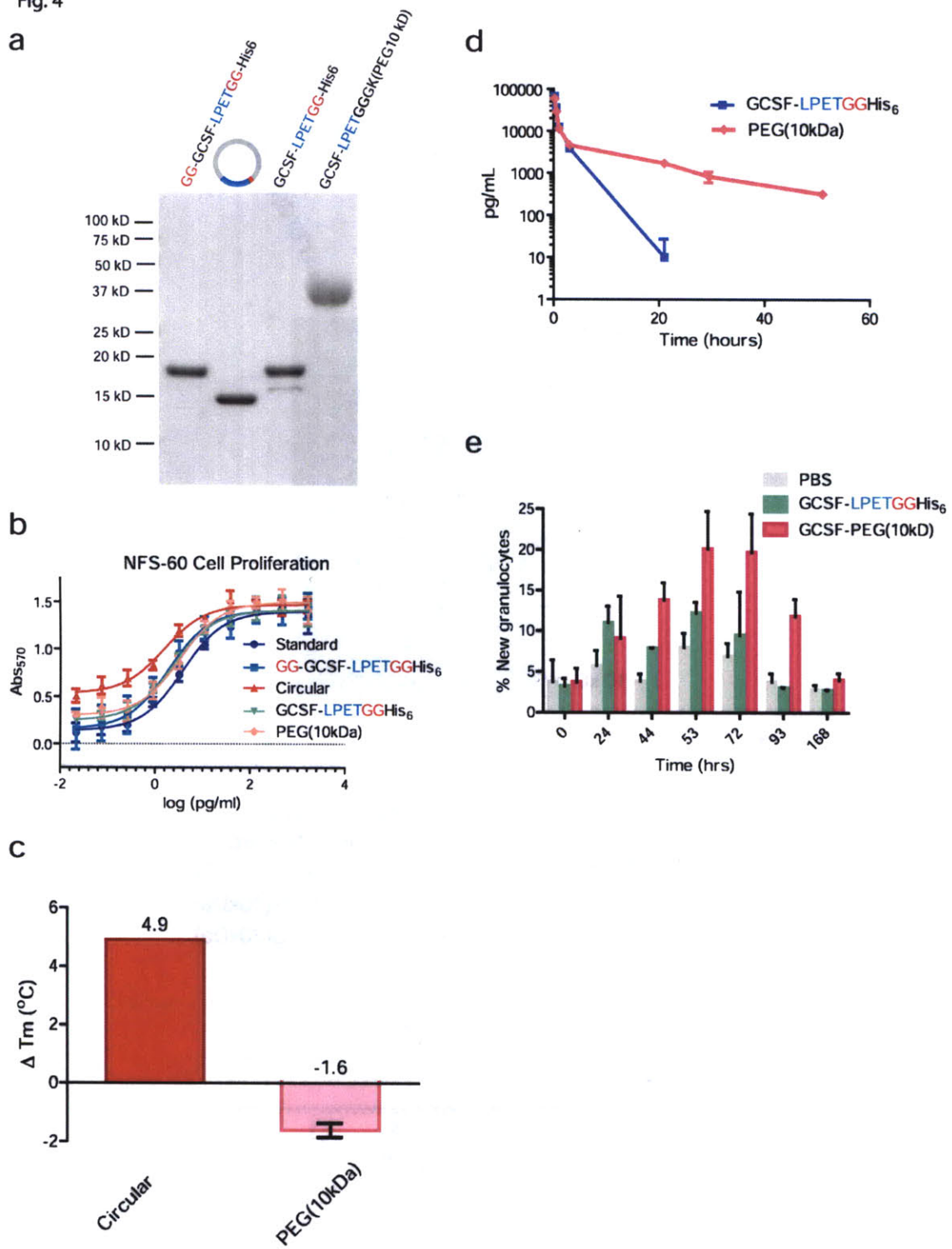
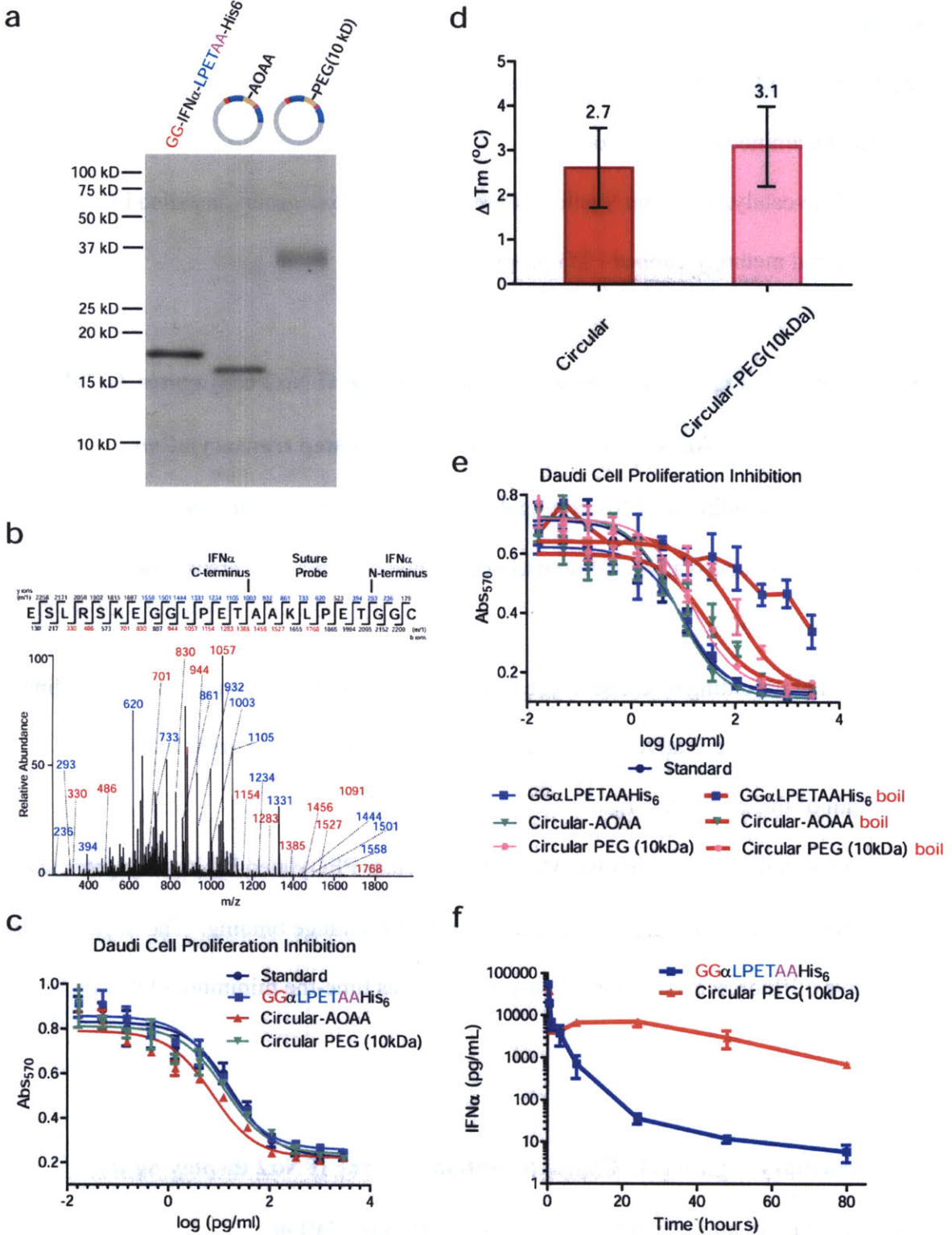


Figure 6.5

Fig. 6



## **Supplementary Figure Legends**

### **Supplementary Figure 6.1. Probes used and aniline catalyzed oxime ligation scheme.**

- (a) Chemical structures of oligoglycine-functionalized PEG probes (top) and SrtA<sub>Strep</sub>-compatible aminoxy probe (bottom).
- (b) Aniline-catalyzed oxime ligation between an aminoxy group, installed by sortase reaction, and methoxy-capped PEG propionaldehyde.

### **Supplementary Figure 6.2. Thermal denaturation of IFN $\alpha$ 2 conjugates, GCSF-3 conjugates, and IFN $\alpha$ 2 conjugates generated by two-step transacylation.**

- (a) Thermal unfolding of IFN $\alpha$  conjugates was monitored by Sypro Orange binding. The negative first derivative of fluorescence change is displayed versus time-the minimum of this plot is the  $T_m$ .
- (b) Thermal unfolding of GCSF-3 conjugates was monitored by Sypro Orange binding. The negative first derivative of fluorescence change is displayed versus time-the minimum of this plot is the  $T_m$ .
- (c) Thermal unfolding of circular-AOAA and circular PEGylated IFN $\alpha$  conjugates, as well as the starting material, was monitored by Sypro Orange binding. The negative first derivative of fluorescence change is displayed versus time-the minimum of this plot is the  $T_m$ .

### **Supplementary Figure 6.3. Characterization of cyclic IFN $\alpha$ 2 displaying the aminoxy group and efficiency of single-step transacylation.**

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

(a) ESI-MS reconstruction of cyclic IFN $\alpha$ 2 displaying the aminoxy group.

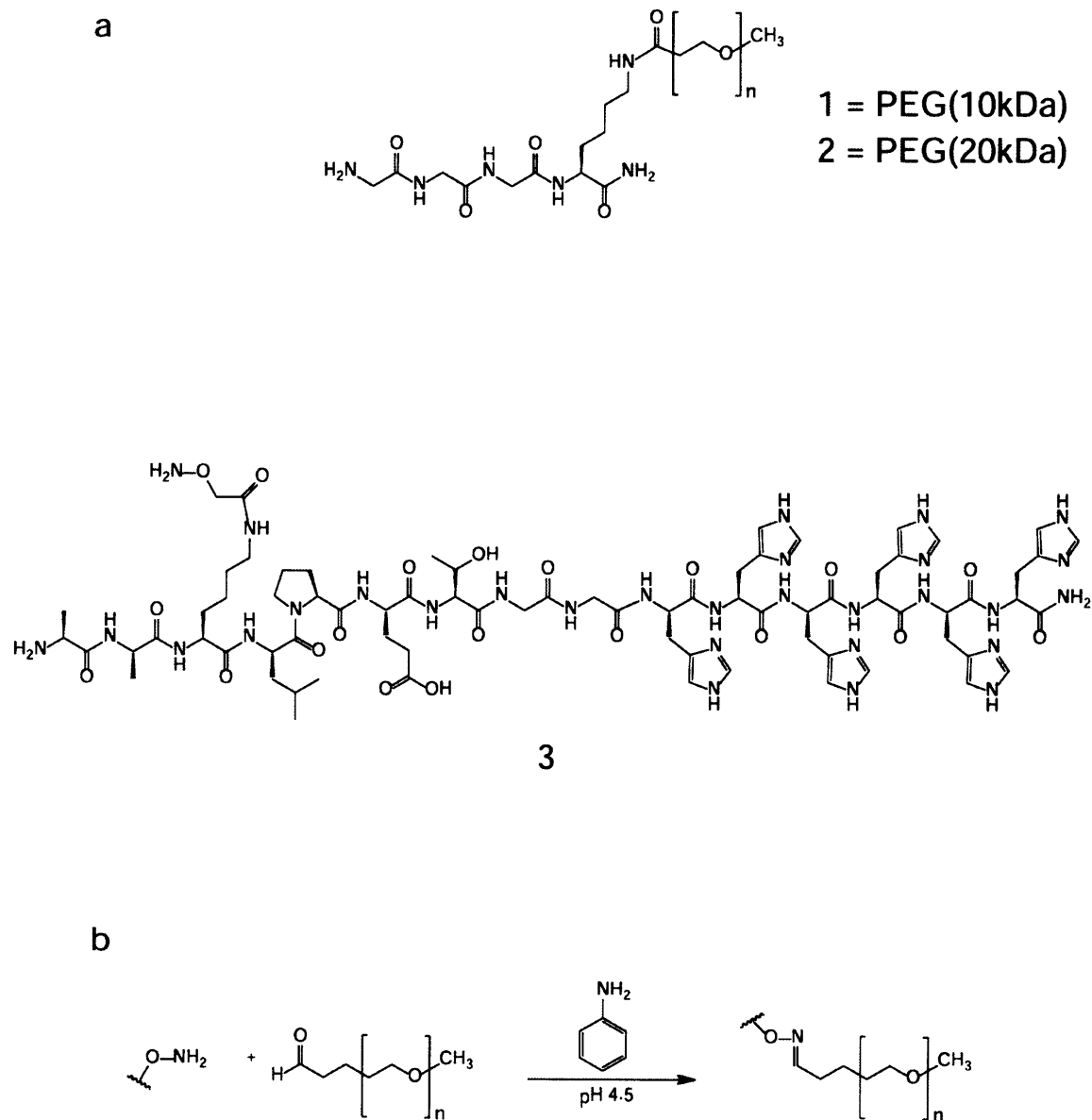
Unprocessed ESI-MS spectrum (inset).

(b) Representative ion exchange chromatography chromatograms of crude sortase reactions for PEGylation of IFN $\alpha$  (left) and GCSF cyclization (right) showing efficient conversion from the input material to the indicated transpeptidation product.

Supplementary Figures

Supplementary Figure 6.1

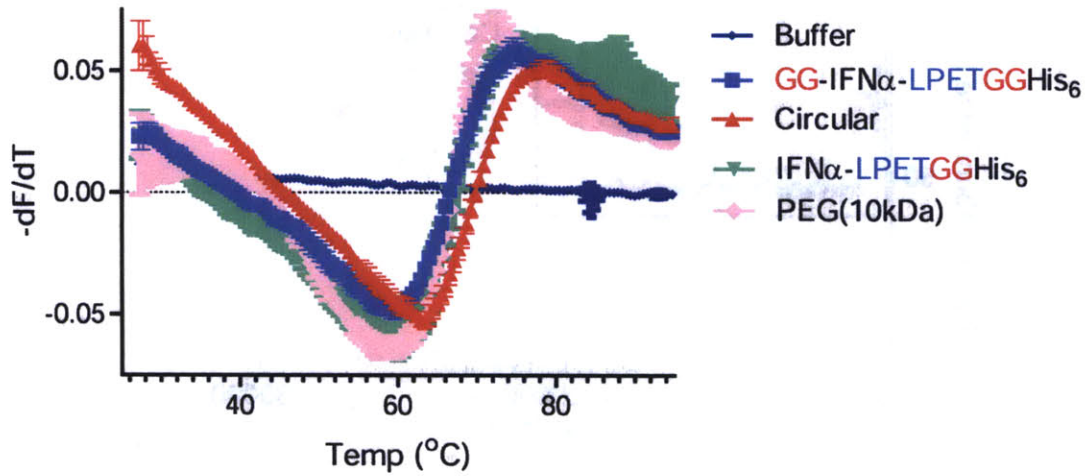
Fig. S1



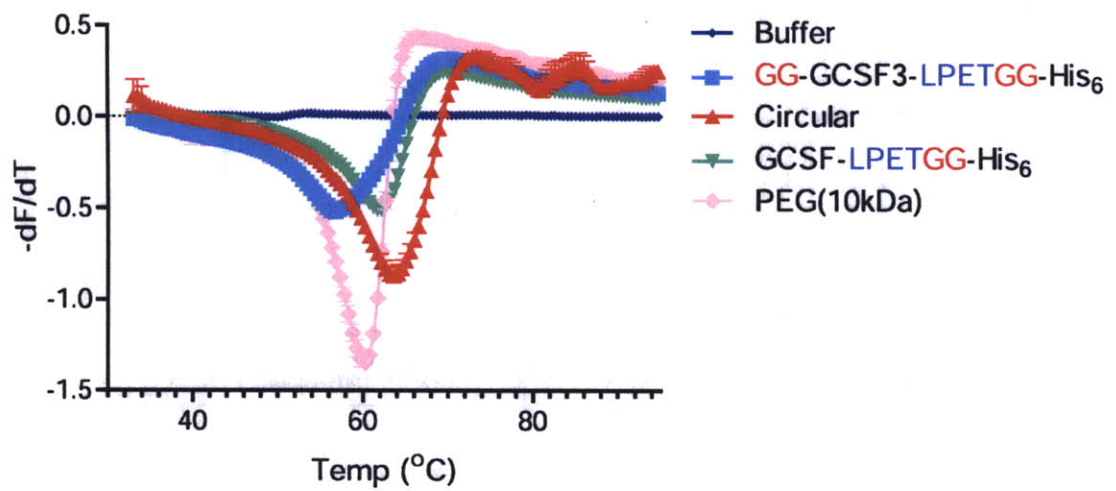
Supplementary Figure 6.2

Fig. S2

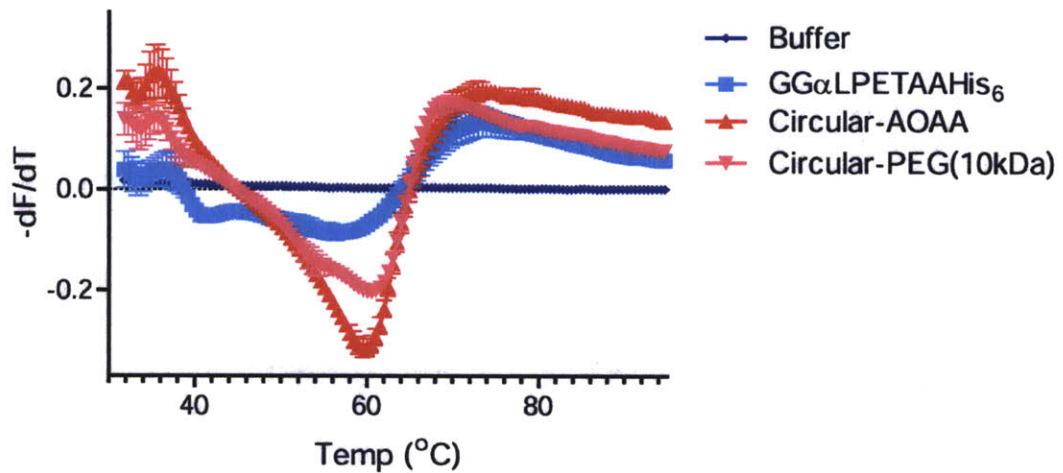
a



b



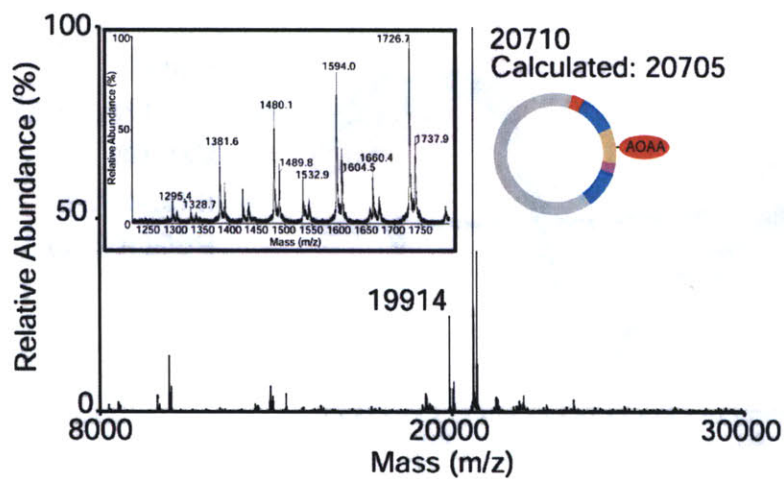
c



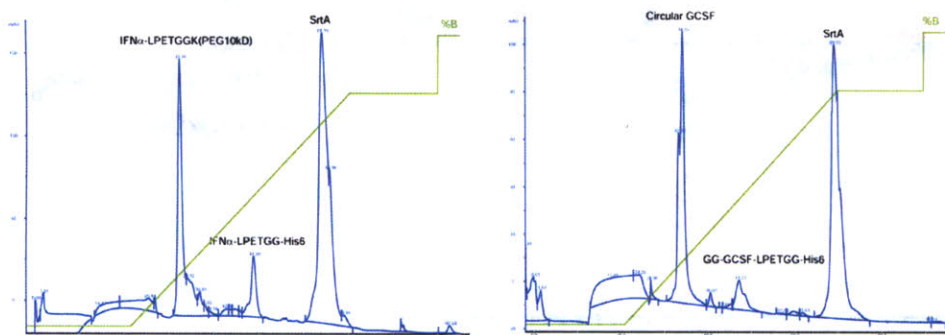
Supplementary Figure 6.3

Fig. S3

a



b





### Supplementary Table 6.1

Table S1. Properties of sortase-modified IFN $\alpha$ 2a.

	$IC_{50}$ (pg/ml)	$T_m$ ( $^{\circ}C$ )	Half-Life (h)
Standard	15.5 $\pm$ 1.1	n.d.	n.d.
GG $\alpha$ LPETGGHis <sub>6</sub>	10 $\pm$ 1.1	59.6 $\pm$ 0	n.d.
Circular	4.9 $\pm$ 1.2	63.2 $\pm$ 0.2	n.d.
$\alpha$ LPETGGHis <sub>6</sub>	7.1 $\pm$ 1.1	59.6 $\pm$ 0.9	$\alpha$ = 0.27, $\beta$ =1.46
PEG(10kDa)	21.7 $\pm$ 1.1	59.2 $\pm$ 1.0	$\alpha$ = 0.32, $\beta$ =5.98

### Supplementary Table 6.2

Table S2. IFN $\alpha$ 2a modified by sortase with different length PEG chains is highly potent.

	$IC_{50}$ (pg/ml)
Standard	33.5 $\pm$ 1.2
$\alpha$ LPETGGHis <sub>6</sub>	35.5 $\pm$ 1.1
PEG(10kDa)	65.34 $\pm$ 1.2
PEG(20kDa)	88.63 $\pm$ 1.2

### Supplementary Table 6.3

Table S3. Properties of sortase-modified GCSF-3.

	$EC_{50}$ (pg/ml)	$T_m$ ( $^{\circ}C$ )	Half-Life (min)
Standard	3.9 $\pm$ 1.1	n.d.	n.d.
GG-GCSF-LPETGGHis <sub>6</sub>	1.9 $\pm$ 1.2	59.6 $\pm$ 0	n.d.
Circular	1.6 $\pm$ 1.2	63.2 $\pm$ 0	n.d.
GCSF-LPETGGHis <sub>6</sub>	2.5 $\pm$ 1.1	59.6 $\pm$ 0.2	$\alpha$ = 0.30, $\beta$ =2.21
PEG(10kDa)	3.5 $\pm$ 1.2	59.2 $\pm$ 0	$\alpha$ = 0.27, $\beta$ =10.59

**Supplementary Table 6.4**

Table S5. Properties of IFN $\alpha$ 2a subjected to two-step transacylation.

	$IC_{50}$ (pg/ml)	$T_m$ ( $^{\circ}C$ )	Half-Life (h)
Standard	17.4 $\pm$ 1.2	n.d.	n.d.
GG $\alpha$ LPETA $\alpha$ His <sub>6</sub>	12.5 $\pm$ 1.2	57.2 $\pm$ 0.9	$\alpha$ = 0.20, $\beta$ =3.53
Circular-AOAA	7.5 $\pm$ 1.2	59.8 $\pm$ 0.3	n.d.
Circular-PEG(10kDa)	13.2 $\pm$ 1.2	60.3 $\pm$ 0.3	$\alpha$ = 0.15, $\beta$ =31.07

**Supplementary Table 6.5**

Table S5. Properties of IFN $\alpha$ 2a subjected to two-step transacylation.

	$IC_{50}$ (pg/ml)	$T_m$ ( $^{\circ}C$ )	Half-Life (h)
Standard	17.4 $\pm$ 1.2	n.d.	n.d.
GG $\alpha$ LPETA $\alpha$ His <sub>6</sub>	12.5 $\pm$ 1.2	57.2 $\pm$ 0.9	$\alpha$ = 0.20, $\beta$ =3.53
Circular-AOAA	7.5 $\pm$ 1.2	59.8 $\pm$ 0.3	n.d.
Circular-PEG(10kDa)	13.2 $\pm$ 1.2	60.3 $\pm$ 0.3	$\alpha$ = 0.15, $\beta$ =31.07

**Supplementary Table 6.6**

TableS6. Properties of IFN $\alpha$ 2a subjected to two-step transacylation.

	<i>IC</i> <sub>50</sub> (pg/ml)
Standard	7.3 $\pm$ 1.1
GG $\alpha$ LPETAHis <sub>6</sub>	11.0 $\pm$ 1.2
GG $\alpha$ LPETAHis <sub>6</sub> boil	–
Circular-AOAA	7.4 $\pm$ 1.1
Circular-AOAA boil	28.8 $\pm$ 1.3
Circular PEG (10kDa)	14.5 $\pm$ 1.1
Circular PEG (10kDa) boil	112.6 $\pm$ 1.3

**Supplementary Text**

**Supplementary Methods**

**Animals.** C57BL/6 mice were purchased from Jackson Labs and used at 6-8 weeks of age. All studies were approved by the MIT Committee on Animal Care.

**Plasmids and Protein Expression.** Human IFN $\alpha$ 2a lacking the leader sequence and fused at the C-terminus to the sequence GGLPETGGHHHHHH was cloned into the pET28a+ vector (Novagen). This same protein was cloned with two glycines at the N-terminus. Human G-CSF3 versions were cloned similarly. For IFN $\alpha$ 2a with the SrtA<sub>Strep</sub> cleavage site, a Quick Change kit (Stratagene) was used to mutate the LPETGG motif to LPETAA. Human erythropoietin was cloned into pLHCX (Clontech), fused at the C-

terminus to GGLPETGGHHHHH. The leader sequence was replaced with the H-2K<sup>b</sup> signal peptide followed by two glycines.

SrtA<sub>Staph</sub> and SrtA<sub>Strep</sub> were expressed and purified as described previously<sup>10, 13</sup>. IFN $\alpha$ 2a variants as well as G-CSF-3 variants were expressed in Rosetta-gami(DE3)pLysS cells (Novagen) with 500  $\mu$ M IPTG (Sigma), grown in 2YT medium for 16 hrs at 25°C. *E.coli* pellets were lysed by sonication in lysis buffer: 50mM Tris, 150 mM NaCl, 10 mM imidazole, 50  $\mu$ g/ml DNaseI (Roche), pH 7.2 and clarified by centrifugation. Soluble protein was purified by Ni-NTA IMAC chromatography (Qiagen) in lysis buffer lacking DNase and eluted with lysis buffer supplemented with 500 mM imidazole. Eluates were diluted with either 150mM MES, 150mM NaCl, pH 5.0 (IFN $\alpha$ 2a variants) or 150mM sodium acetate/150mM NaCl, pH4.0 (G-CSF3 variants). Proteins were further purified by size exclusion chromatography on a Superdex 75 column (GE) using either 20mM MES, 150mM NaCl, pH 5.0 (IFN $\alpha$ 2a) or 20 mM sodium acetate, 150 mM NaCl, pH 4.0 (G-CSF3) as eluent. Protein concentration was determined by the Bradford method.

Erythropoietin variants were expressed in HEK-293T cells. Retrovirus was produced as described previously<sup>32</sup> and used to infect HEK-293T cells. Stably transduced cells were cultured in DME/10%IFS/ 50 units/ml penicillin, 50  $\mu$ g/ml streptomycin sulfate, and 0.125  $\mu$ g/ml amphotericin B (Fungizone) supplemented with 0.125 mg/ml Hygromycin B (Roche). For protein purification, cells were cultured to ~80% confluency in complete medium. Medium was harvested and supplemented with 1/10 volume of a 10X buffer containing 500 mM Tris, pH 8.0, 1500 mM NaCl, 100 mM Imidazole and 500 mL of this

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

conditioned media was incubated for 2 hours at 25°C with 1 mL of Ni-NTA (Qiagen). Material was loaded into disposable plastic columns (Biorad) and washed using 50mM Tris, 150 mM NaCl, 10 mM imidazole, pH 7.2. Protein was eluted with wash buffer supplemented with 500 mM imidazole and used without further desalting/purification.

**Nucleophile Preparation.** The Gly<sub>3</sub>K peptide scaffold was constructed by standard Fluorenylmethoxycarbonyl (Fmoc) solid phase peptide chemistry on rink amide resin. Fmoc protected peptide was liberated from the resin by treatment with 95% TFA/3%TIPS/2%H<sub>2</sub>O, precipitated with cold ether and lyophilized. Methoxy-capped 10 kDa PEG succinimidyl ester (1 equivalent, Nanocs) or 20 kDa PEG succinimidyl ester was mixed with peptide (2equivalents), 1 equivalent N-hydroxysuccinimide (NHS, Acros Organics) and 1 equivalent 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC, Pierce) in *N*-Methylpyrrolidone (NMP, Fluka) for 24 hours at room temperature, followed by precipitation in cold ether. The resulting solid was resuspended in 20% piperidine/NMP for 30 minutes, followed by re-precipitation in cold ether. This material was resuspended in H<sub>2</sub>O and dialyzed extensively against H<sub>2</sub>O to remove free peptide and afford Gly<sub>3</sub>K-PEG(10kDa) and Gly<sub>3</sub>K-PEG(20kDa) nucleophiles compatible with SrtA<sub>staph</sub> labeling.

The AAKLPETGGHHHHHH peptide scaffold was synthesized on Rink Amide resin by the MIT Biopolymers Facility using lysine(MTT) in position 3. Fmoc protected peptide was MTT-deprotected on-resin after swelling in dichloromethane (DCM, JT Baker) and treatment with 95% DCM/3%TIPS/2%TFA. Deprotected resin-bound peptide was incubated with 5 equivalents Bis-Boc-Aminoxyacetic acid (AOAA, Novagen), 5

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

equivalents N,N'-diisopropylcarbodiimide (American Bioanalytical), 5 equivalents NHS, and 5 equivalents anhydrous N-Hydroxybenzotriazole (HoBt, American Bioanalytical) in Dimethylformamide (Sigma) overnight with agitation. Resin was washed extensively with DMF and NMP and then treated with 20% piperidine/NMP for 30 minutes to remove the N-terminal Fmoc group. Next, resin was washed extensively in DCM then the peptide was liberated with 95% TFA/3%TIPS/2%H<sub>2</sub>O for 3 hours, precipitated in ice-cold ether and lyophilized to afford H-AAK(AOAA)LPETGGHHHHHH-NH<sub>2</sub> peptide. Peptide was characterized by LC-MS: [M+2H]<sup>2+</sup>=869.9, obs=869.2, [M+3H]<sup>3+</sup>=580.3, obs=580.1 and used without further purification.

**Mass Spectrometry.** LC/MS analysis was performed using a Micromass LCT mass spectrometer and an Agilent 1100 Series HPLC system equipped with a Waters Symmetry 3.5 μM C18 column (2.1 x 50 mm, MeCN:ddH<sub>2</sub>O gradient mobile phase containing 0.1% formic acid, 150 μL/min).

For MS/MS analysis, proteins were resolved on 12.5% SDS PAGE gels, stained with coomassie, and the relevant bands were excised, destained, and subjected to trypsinolysis (circular IFNα2a, circular EPO) or AspN (circular IFNα2a-AOAA). Recovered peptides were analyzed by reversed-phase liquid chromatography electrospray ionization mass spectrometry using a Waters nanoACQUITY-UPLC coupled to a Thermo LTQ linear ion-trap mass spectrometer. MS/MS spectra were searched against a custom database with circularly permuted sequences for each protein using SEQUEST. SEQUEST results

Chapter 6: Sortase-catalyzed transformations that improve the properties of cytokines

were analyzed with Bioworks Browser 3.3 and filtered with the following criteria:

different peptides,

minimum cross correlation coefficients (1, 2, 3 charge states) of 1.50, 2.00, 2.50, number

different peptides of 2 per protein and Sp – preliminary score of 300.



Print Article (reprinted with permission from Proceedings of the National Academy of Sciences):

PNAS

PNAS

PNAS

PNAS

PNAS

PNAS

PNAS

PNAS

PNAS

PNAS

PNAS

PNAS

## Sortase-catalyzed transformations that improve the properties of cytokines

Maximilian W. Popp<sup>a,b</sup>, Stephanie K. Dougan<sup>a</sup>, Tzu-Ying Chuang<sup>a</sup>, Eric Spooner<sup>a</sup>, and Hidde L. Ploegh<sup>a,b,1</sup>

<sup>a</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142; and <sup>b</sup>Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02142

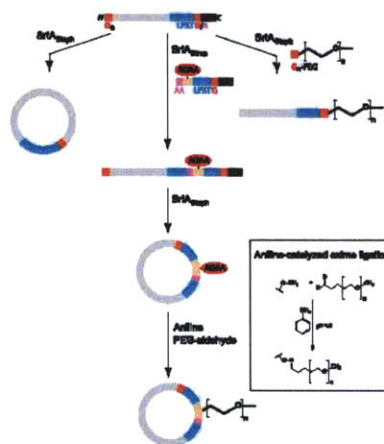
Edited by Donald F. Steiner, University of Chicago, Chicago, IL, and approved January 3, 2011 (received for review November 9, 2010)

Recombinant protein therapeutics often suffer from short circulating half-life and poor stability, necessitating multiple injections and resulting in limited shelf-life. Conjugation to polyethylene glycol chains (PEG) extends the circulatory half-life of many proteins, but the methods for attachment often lack specificity, resulting in loss of biological activity. Using four-helix bundle cytokines as an example, we present a general platform that uses sortase-mediated transpeptidation to facilitate site-specific attachment of PEG to extend cytokine half-life with full retention of biological activity. Covalently joining the N and C termini of proteins to obtain circular polypeptides, again executed using sortase, increases thermal stability. We combined both PEGylation and circularization by exploiting two distinct sortase enzymes and the use of a molecular suture that allows both site-specific PEGylation and covalent closure. The method developed is general, uses a set of easily accessible reagents, and should be applicable to a wide variety of proteins, provided that their termini are not involved in receptor binding or function.

Many clinically relevant cytokines share a four-helix bundle structure, typified by IFN $\alpha$ 2, Granulocyte colony-stimulating factor 3 (G-CSF-3), Erythropoietin (EPO), IL-2, IL-4, IL-7, IL-9, and IL-15. Cocystal structures of cytokines with receptor fragments and biochemical studies that map residues critical for interaction of a cytokine with its receptor show that the receptor contacts the sides of the helical bundles (1). This mode of interaction positions the N and C termini of the cytokine away from the receptor.

Polyethylene glycol chains attached to therapeutically important proteins increase circulatory half-life, reduce clearance by kidney filtration, reduce proteolysis, and reduce the generation of neutralizing antibodies (2–4). The attachment of PEG commonly employs standard chemistries that target reactive amino acid side chains (e.g., cysteine and lysine). This strategy often generates a heterogeneous mixture in which multiple amino acids in the target are modified with a PEG chain, necessitating cumbersome separations and characterization (5). Such PEGylated molecules often show decreased biological activity, likely due to attachment of a PEG chain to a residue important for interaction with the receptor (6)—this problem may be overcome by site-specific PEGylation. Although engineering of a carefully placed unpaired cysteine residue allows site-specific PEGylation (7, 8), this method must be tailored to the specific protein target.

Sortase A from *Staphylococcus aureus* (SrtA<sub>staph</sub>) is a thiol-containing transpeptidase that recognizes an LPXTG motif in multiple structurally unrelated substrates (9). SrtA<sub>staph</sub> cleaves the peptide bond between the threonine and glycine residues with concomitant formation of a thioacyl enzyme intermediate that involves the catalytic cysteine and the substrate threonine. This acyl-enzyme is resolved by nucleophilic attack by the N terminus of an oligoglycine peptide, resulting in formation of an amide bond between the substrate protein and the incoming nucleophile (10). Sortase A tolerates C-terminal extensions of the oligoglycine nucleophile, allowing diverse functionalized nucleophiles to be installed site specifically onto proteins equipped with an LPXTG motif (11, 12). The related *Streptococcus pyogenes*



**Fig. 1.** The sortase reaction scheme. Cyclization of substrates equipped with a C-terminal LPXTG sortase recognition element as well as an N-terminal glycine by SrtA<sub>staph</sub> (left). Substrates lacking the N-terminal glycine and incubated in the presence of exogenous oligo-glycine functionalized PEG are site-specifically PEGylated (right). Circular, PEGylated proteins can be constructed by incubating substrates bearing an N-terminal glycine and a C-terminal SrtA<sub>staph</sub> cleavage site (LPXTA) with an alanine-based nucleophile carrying an amino oxime group and the SrtA<sub>staph</sub> cleavage site (Center). This transpeptidation product is cyclized upon incubation with SrtA<sub>staph</sub>. Finally, PEG is attached by aniline-catalyzed oxime ligation with methoxy-capped PEG proprionaldehyde. The oxime bond formed is shown (inset).

sortase accepts di-alanine based nucleophiles, which the *S. aureus* enzyme does not. This sortase (SrtA<sub>strep</sub>) cleaves the LPXTA motif between threonine and alanine and allows installation of modified alanine-based nucleophiles. SrtA<sub>strep</sub> also recognizes and cleaves LPXTG motifs, albeit with reduced efficiency, however the LPXTA motif is refractory to cleavage by SrtA<sub>staph</sub> (13, 14).

Here we present a general strategy for site-specific modification of therapeutic recombinant proteins at their termini, an approach that is particularly well suited to the four-helix bundle cytokines, where the termini are distant from the receptor interaction site. We show improvements in thermal stability by cova-

Author contributions: M.W.P. and H.L.P. designed research; M.W.P., S.K.D., T.-Y.C., and E.S. performed research; M.W.P., S.K.D., and E.S. analyzed data; and M.W.P. and H.L.P. wrote the paper.

Conflict of interest statement: The authors have applied for a patent based on the technology used in the paper.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

To whom correspondence should be addressed. E-mail: ploegh@wi.mit.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015863108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015863108/-DCSupplemental).



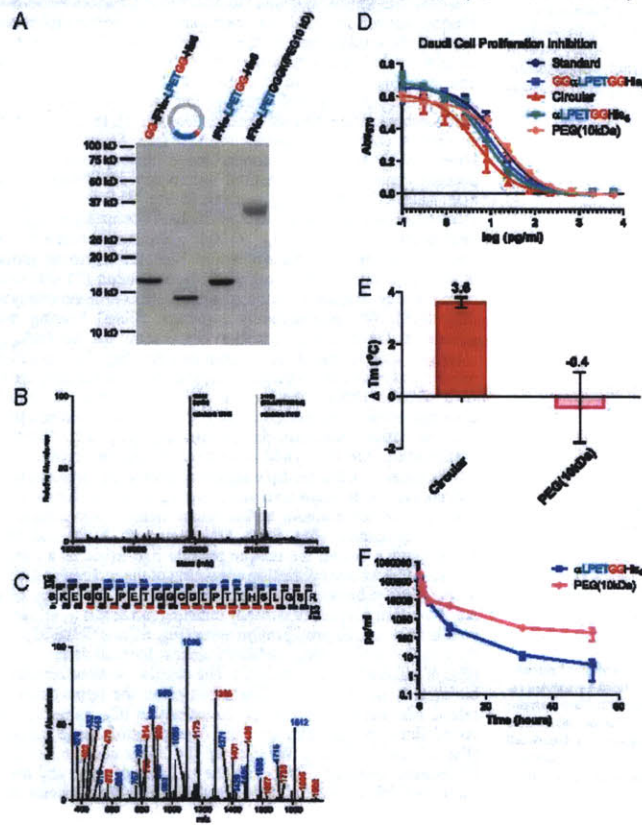
lently joining the N and C termini of these proteins (15). We present a general method for combining both site-specific PEGylation and improved thermal stability in a single molecule by using a molecular suture produced in two successive rounds of sortase-catalyzed transpeptidation (Fig. 1).

**Results**

**Modification of IFN $\alpha$ 2.** We first applied the sortase reaction to human interferon alpha (IFN $\alpha$ 2), a four-helix bundle cytokine where chemical PEGylation results in multiple positional isomers and a ~90% decrease in biological potency (16). We expressed in *Escherichia coli* two versions of IFN $\alpha$ 2, both with an LPETG sortase motif immediately followed by a hexahistidine tag for ease of purification. At the amino terminus, one of the two proteins possesses a diglycine motif for cyclization. This material can either be cyclized or PEGylated at the C terminus (Fig. 2A). IFN $\alpha$ 2 with an N-terminal diglycine motif and a C-terminal LPETG was successfully cyclized by incubation with SrtA<sub>Staph</sub> and purified to homogeneity by ion-exchange chromatography. The mass of the circularized product in a crude sortase reaction differs from a linear hydrolysis product by -18 Da, consistent with an intramolecular transpeptidation reaction (Fig. 2B). We identified the unique junction peptide that arises from intramolecular cyclization by tandem mass spectrometry (MS/MS) analysis of tryptic digests (Fig. 2C). Site-specific PEGylation of the IFN $\alpha$ 2 variant that lacks the N-terminal diglycine motif (and

thus unable to undergo intramolecular cyclization) was achieved by incubation with sortase and a GGGK peptide to which a 10 kDa PEG moiety was affixed via the lysine  $\epsilon$ -amine (Probe 1, Fig. 2A and Fig. S1A). We confirmed equipotent biological activity of all four interferon alpha variants in a Daudi cell proliferation inhibition assay when compared to a commercially available, non-PEGylated preparation (Fig. 2D and Table S1). Addition of a protein-sized PEG module to the C terminus of IFN $\alpha$ 2 does not appreciably perturb receptor binding and biological potency. We hypothesized that the circular IFN $\alpha$ 2 would be more resistant to thermal denaturation, because the termini are clamped shut and thus should not “fray” (15, 17). We measured thermal denaturation with a ThermoFluor assay, using the commercially available Sypro Orange dye (18). Indeed, the circular form of IFN $\alpha$ 2 had a significantly elevated melting temperature ( $T_m$ ) compared to the other three variants, all of which exhibited nearly identical  $T_m$ 's (Fig. 2E, Table S1, and Fig. S2A). Finally, we injected mice with either the linear IFN $\alpha$ 2 variant used to construct the PEGylated form, or the IFN $\alpha$ 2 variant bearing a 10 kDa PEG chain (Probe 1) at its C terminus and measured the serum levels over time by ELISA. The PEGylated form decayed with slower kinetics than its non-PEGylated counterpart (Fig. 2F and Table S1).

**Modification of a Single Protein Preparation by Multiple Nucleophiles.** The sortase platform offers the ability to install nearly any



**Fig. 2.** Synthesis and characterization of human IFN $\alpha$ 2 conjugates. (A) GG-IFN $\alpha$ -LPETGGH<sub>6</sub> was used to obtain backbone-cyclized IFN $\alpha$ 2, and IFN $\alpha$ -LEPTGGH<sub>6</sub> was used to make IFN $\alpha$ -PEG (10 kDa), bearing a C-terminal PEG chain. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie. (B) ESI-MS characterization of IFN $\alpha$  species in crude sortase reactions either before (linear GG-IFN $\alpha$ -LPETGGH<sub>6</sub>) or after (cyclic IFN $\alpha$ 2) overnight incubation at 25 °C. (C) MS/MS identification of a tryptic peptide comprising the IFN $\alpha$ 2 C terminus, followed by the SrtA<sub>Staph</sub> cleavage site, joined to the N-terminal glycines in the IFN $\alpha$ 2 precursor. (D) In vitro bioactivity of IFN $\alpha$  conjugates in a Daudi cell proliferation inhibition assay. Dose-response curves for Daudi cell proliferation inhibition were measured by [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium] bromide (MTT) assay, in triplicate, with the standard deviation displayed. (E) Resistance of backbone-cyclized IFN $\alpha$  to thermal denaturation. Change in  $T_m$  for circular and PEGylated IFN $\alpha$  conjugates, relative to their precursor proteins, were measured by thermal shift assay, with  $n = 4$ . (F) Change in serum IFN $\alpha$  levels following tail vein injection of site-specifically PEGylated IFN $\alpha$  as well as its precursor protein. 10  $\mu$ g of each protein was injected and, at the indicated time points, blood was withdrawn and cytokine concentration in serum was determined by ELISA assay. Data are means  $\pm$  standard deviation for three mice in each group.



nongenetically encoded entity onto the C terminus of a single preparation of recombinant protein. We demonstrate this ability by synthesizing a nucleophile that bears a 20 kDa PEG moiety (Probe 2 and Fig. S1A), as well as the previously described 10 kDa PEG probe (Probe 1), and appended them to the identical preparation of IFN $\alpha$ 2 (Fig. 3A). Both conjugates were tested for biological activity (Fig. 3B) and are highly potent (Table S2).

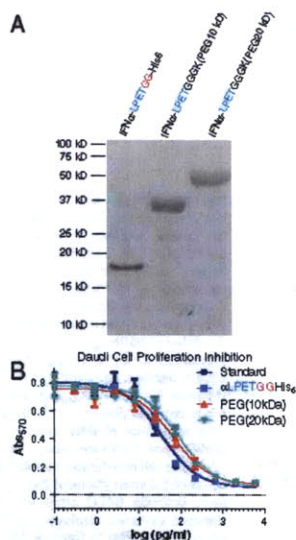
**Superior Properties Endowed upon Sortase Substrates Are Applicable to Multiple Proteins.** Sortase allows the installation of the same nonnatural group onto different proteins, unrelated in sequence or amino acid composition. Accordingly, we applied the sortase reaction to a different four-helix bundle cytokine, GCSF-3, with similar results. We first made the linear precursor GCSF variants and then cyclized or PEGylated them with probe 1 using sortase (Fig. 4A). All variants possess equal or superior in vitro bioactivity relative to a commercial non-PEGylated preparation, as assessed by a standard cell proliferation assay with the NFS-60 murine myeloblastic cell line (19) (Fig. 4B and Table S3). Like cyclic IFN $\alpha$ 2, cyclic GCSF-3 shows increased thermostability (Fig. 4C, Fig. S2B, and Table S3) and PEGylated GCSF-3 has a significantly extended circulatory half-life relative to its non-PEGylated precursor (Fig. 4D and Table S3) upon injection into mice. PEGylated GCSF-3 also led to a more robust and prolonged proliferation of granulocytes in injected mice than its non-PEGylated analog (Fig. 4E).

**Modification of Glycoproteins.** Many important therapeutics are glycosylated proteins that traverse the secretory pathway and are produced in mammalian cell culture. Are such proteins equally amenable to engineering using a sortase-mediated transpeptidation reaction? We chose human erythropoietin (EPO) because

~40% of its mass is composed of bulky, charged N-linked glycans, and cyclization provides a stringent test of the sortase method, as it requires (minimal) modifications at both termini. We replaced the endogenous EPO signal sequence with the murine H-2K<sup>b</sup> signal sequence, followed by two glycine residues. This signal peptide is cleaved in mammalian cells with concomitant exposure of an N-terminal glycine residue. We added the requisite sortase motif followed by a hexahistidine tag at the C terminus of EPO for ease of purification. Human embryonic kidney (HEK)-293T cell lines stably transduced with this construct yielded preparations that were <50% pure after a single round of Nickel-nitrilotriacetic acid immobilized metal affinity chromatography (Ni-NTA IMAC) of conditioned medium. Material from such preparations can be cyclized successfully upon incubation with sortase (Fig. 5A). Again, we identified the unique junction peptide that arises from covalent ligation of the N and C termini by MS/MS analysis of cyclized EPO, shown of all glycans by incubation with peptide: N-glycosidase F (PNGaseF), and digested with trypsin (Fig. 5B). Both the linear preparations as well as the sortase reaction containing cyclic EPO were tested in a cell proliferation assay with Ba/F3 cells that stably express the erythropoietin receptor (20) and these compared favorably with commercial EPO preparations. Most importantly, the cyclic form is as potent as its linear counterpart (Fig. 5C and Table S4), indicating that modification of the EPO termini does not affect its biological activity.

**Two-Step Transacylation Allows Combination of Enhanced Properties.**

Having shown for multiple examples that sortase-mediated site-specific modification of the C terminus with PEG increases circulatory half-life with nearly no loss of biological activity, and that covalent closure of the N and C termini yield proteins stabilized against thermal denaturation, we executed a scheme that combines these desirable properties (Fig. 1). In the first step, we exploit a previously described sortase from *Streptococcus pyogenes* (13, 14) that accepts alanine-based nucleophiles to affix a peptide containing a nonnatural amino oxy containing amino acid, followed by the SrtA<sub>Staph</sub> cleavage site. We use SrtA<sub>Staph</sub> to effectuate covalent closure, as the LPETAA site left by SrtA<sub>Staph</sub> is resistant to attack by SrtA<sub>Staph</sub>, to yield a circular protein with the amino oxy group at the place of suture. Next, the amino oxy group is used in a bioorthogonal oxime ligation reaction (21–23) with commercially available methoxy-capped PEG-propionaldehyde (Fig. S1B). We recombinantly expressed IFN $\alpha$ 2 bearing two glycines at the N terminus, and at the C terminus the SrtA<sub>Staph</sub> cleavage site, followed by a hexahistidine tag. The purified protein was subjected to the two-step transacylation procedure to yield circular IFN $\alpha$ 2 bearing the amino oxy group. After ion exchange chromatography, this material was PEGylated and purified by cation exchange chromatography (Fig. 6A). Without optimization, the final yield of circular PEGylated protein was ~10%, because of the multiple separate chromatography and protein concentration steps involved in production and PEGylation. The mass of the material before oxime ligation was consistent with its cyclization (Fig. S3A). Digestion with AspN and MS/MS analysis revealed the unique peptide that arises as a consequence of backbone cyclization consisting of the molecular suture probe stitched between the C and N termini of IFN $\alpha$  (Fig. 6B). All three variants yielded virtually indistinguishable IC<sub>50</sub> values in an inhibition of cell proliferation assay (Fig. 6C and Table S5) and the circular versions were stabilized against thermal denaturation (Fig. 6D, Table S5, and Fig. S2C). The circular versions remained biologically potent after boiling when given the opportunity to refold, whereas we were unable to extract an IC<sub>50</sub> value for the boiled linear preparation over the concentration range assayed (Fig. 6E and Table S6). We injected the linear and the circular, PEGylated forms of IFN $\alpha$ 2 into the tail vein of mice and measured half life by ELISA. The circular, PEGylated species was

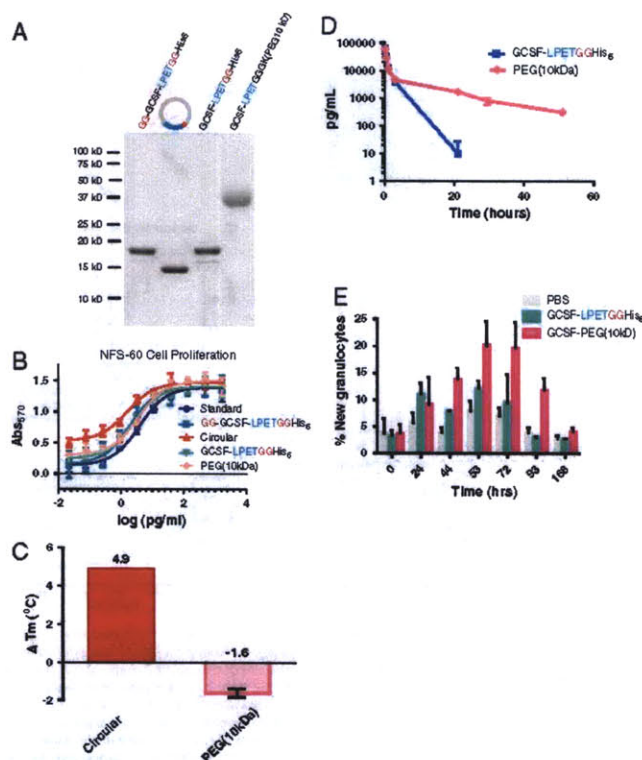


**Fig. 3.** Sortase installs multiple probes onto a single preparation of protein. (A) The same preparation of IFN $\alpha$ -LEPTGGH<sub>6</sub> was PEGylated by sortase reaction with either a 10 kDa or a 20 kDa PEG nucleophile. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie. (B) In vitro bioactivity of 10 kDa and 20 kDa PEGylated IFN $\alpha$  conjugates in a Daudi cell proliferation inhibition assay. Dose-response curves for Daudi cell proliferation inhibition were measured by MTT assay, in triplicate, with the standard deviation displayed.

Popp et al.

PNAS | February 22, 2011 | vol. 108 | no. 8 | 3171





**Fig. 4.** Synthesis and characterization of human GCSF-3 conjugates. (A) GG-GCSF-LPETGGHIS<sub>6</sub> was used to make backbone cyclized GCSF-3, and GCSF-LEPTGGHIS<sub>6</sub> was used to make GCSF-PEG (10 kDa), bearing a C-terminal PEG chain. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie. (B) In vitro bioactivity of GCSF-3 conjugates in an NFS-60 cell proliferation assay. Dose-response curves for NFS-60 cell proliferation were measured by MTT assay, in triplicate, with the standard deviation displayed. (C) Resistance of backbone cyclized GCSF-3 to thermal denaturation. Change in  $T_m$  for circular and PEGylated GCSF-3 conjugates, relative to their precursor proteins, were measured by thermal shift assay ( $n = 4$ ). (D) Change in serum GCSF levels following tail vein injection of site-specifically PEGylated GCSF-3, as well as its precursor. 5  $\mu$ g of each protein was injected and, at the indicated time points, blood was withdrawn and cytokine concentration in serum was determined by ELISA assay. Data are means from three mice in each group, with the standard deviation displayed. (E) GCSF, PEGylated via sortase, is highly potent in vivo. C57BL/6 mice ( $n = 4$  per group) were injected intravenously with linear GCSF, GCSF-PEG (10 kDa), or saline control (PBS). Peripheral blood was collected at the indicated times and stained with antibodies to Gr-1 and CD11b. Newly generated granulocytes were defined as Gr-1 intermediate CD11b<sup>+</sup> cells (31).

cleared significantly more slowly than the linear form (Fig. 6F and Table S5).

#### Discussion

The sortase transpeptidase reaction allows facile site-specific PEGylation of multiple distinct proteins. In all cases tested, the site-specific C-terminal PEGylation proceeds efficiently (Fig. S3B) and yields adducts of known stoichiometry that are biologically equipotent to the non-PEGylated versions, but retain the increase in circulatory half life associated with PEG modification. We attribute these properties to the fact that all enzymatic transformations are performed on native proteins, procedures that should not affect overall conformation or exposure of functionally important side chains. Because our approach to PEGylation is site specific, all preparations are also homogenous and easily purified (5). Although we limit our examples to four-helix bundle cytokines, this platform should be readily extended to structurally distinct therapeutic proteins, with the singular requirement that the C terminus is not involved in receptor binding. A sortase-based approach requires the genetic fusion of a very small (five amino acid) tag to the protein of interest and all transformations occur under native conditions. The additional amino acids that result from fusion to the sortase recognition motif are close to the site of PEG attachment. Immunogenicity of this site is very likely reduced by the PEG moiety shielding this area (3). We and others have labeled proteins with an exposed N-terminal glycine using sortase (14, 24). This approach should therefore be readily extended to site-specific PEGylation at the N terminus as well.

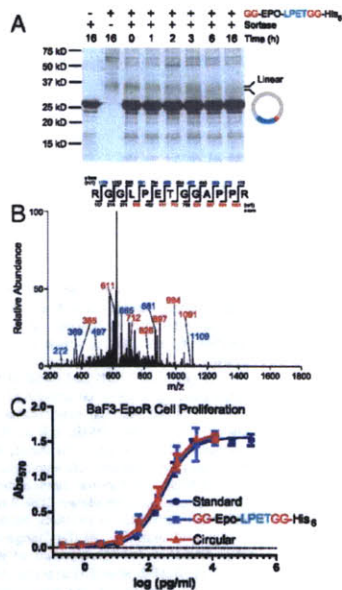
Cyclization of the backbone of protein-based therapeutics also proceeds efficiently (Fig. S3B) and yields preparations that are more resistant to thermal denaturation. Cyclic proteins are also resistant to exoproteolytic attack (25, 26), a feature that may enhance utility of any therapeutic proteins exposed to exoproteases, for example upon receptor mediated internalization. In addition, cyclization of proteins and peptides has been shown to improve potency, stability, and oral bioavailability (27–30).

Finally, we have inserted a nontemplate encoded entity in what is topologically internal to a circular protein, a feat that cannot be accomplished genetically by intein-based methods or by any other currently known means. More generally, this dual transacylation scheme can be used to insert nonnatural groups between fully native proteins expressed separately, or between protein domains. We suspect that this approach will find application in the protein engineering field and so further extend the utility of sortases as tools.

#### Experimental Procedures

**Sortase Reactions.** Sortase reactions with SrtA<sub>Staph</sub> were performed as described previously (11). For either cyclization or PEGylation, reactions containing 50  $\mu$ M substrate, 50  $\mu$ M SrtA<sub>Staph</sub> and 1 mM probe (for PEGylation) were incubated overnight at 25 °C without agitation. Reactions were purified by cation exchange chromatography on a Mono-S (GE Healthcare). For IFN $\alpha$ 2 reactions, continuous gradient chromatography was performed with 20 mM 2-(N-morpholino) ethane sulfonic acid (MES) pH 5.0 and 20 mM MES, 1 M NaCl pH5.0, as eluent. For GCSF reactions, continuous gradient chromatography was performed

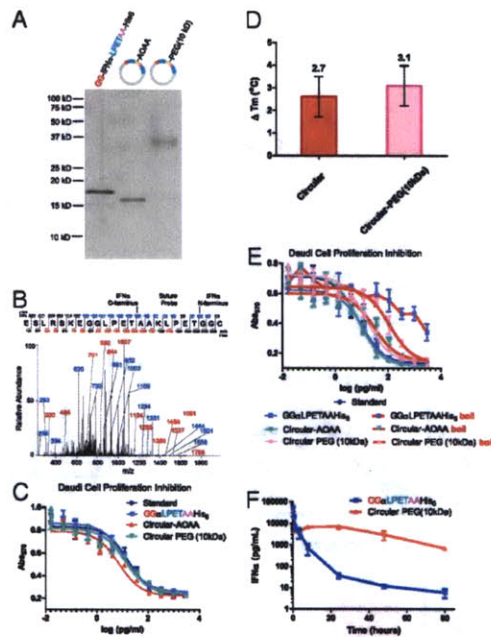




**Fig. 5.** Synthesis and characterization of cyclic human erythropoietin. (A) GG-EPO-LPETGGH<sub>5c</sub> bearing the H-2Kb signal sequence was expressed in HEK-293T cells. Conditioned media was subjected to Ni-NTA IMAC purification. Eluted material was supplemented with 100  $\mu$ M SrtA<sub>Staph</sub> and incubated at 25  $^{\circ}$ C. Aliquots were removed at the indicated time points and subjected to 12.5% SDS-PAGE and protein was visualized by silver staining. (B) MS/MS identification of a peptide generated by tryptic digestion containing the C terminus of EPO, followed by the SrtA<sub>Staph</sub> cleavage site and joined to the N terminus of EPO. (C) *In vitro* bioactivity of linear and cyclic EPO conjugates in a cell proliferation assay using BaF3 cells expressing the erythropoietin receptor (BaF3-EPOR). Eluted material from Ni-NTA IMAC purification was incubated in the presence or absence of sortase for 16 h at 25  $^{\circ}$ C and these crude reactions were measured for *in vitro* EPO bioactivity. Dose response curves for BaF3-EPOR cell proliferation were measured by MTT assay, in triplicate, with the standard deviation displayed. Concentrations were not adjusted for impurities in the EPO preparations.

with 20 mM sodium acetate pH 4.5 and 20 mM sodium acetate, 1 M NaCl pH 4.5, as eluent. Peaks containing the desired product were pooled, concentrated with Vivaspin 500 centrifugal concentrators (Sigma) and protein concentration was determined by Bradford Assay (BioRad). For EPO cyclization, 300  $\mu$ L of Ni-NTA eluate (0.101 mg/mL total protein) was incubated with 100  $\mu$ M SrtA and sortase buffer in 400  $\mu$ L total volume at 25  $^{\circ}$ C for 16 h. For two-step transacylation reactions, 150  $\mu$ M substrate was incubated with 50  $\mu$ M SrtA<sub>strep</sub> and 2 mM probe in 100 mM Tris, 150 mM NaCl pH 8.0 for 4 h at 37  $^{\circ}$ C. SrtA<sub>strep</sub> activity was halted by incubation with 500  $\mu$ M of the protease inhibitor E-64 on ice for 1 h. Crude reaction mixtures were then supplemented with 10 mM imidazole and subjected to Ni-NTA chromatography (Qiagen) to remove prematurely cyclized material. Protein was eluted with 50 mM Tris, 150 mM NaCl, 500 mM imidazole, and buffer exchanged into SrtA<sub>Staph</sub> buffer (50 mM Tris, 150 mM NaCl, and 10 mM CaCl<sub>2</sub>) with a PD10 desalting column (GE Healthcare). For cyclization, this material was incubated overnight with 50  $\mu$ M SrtA<sub>Staph</sub> and purified by cation exchange as described for other IFN $\alpha$ 2a reactions. Concentrated, purified circular IFN $\alpha$ 2 bearing the Aminoxyacetic acid (AOAA) group was then diluted with 1 volume of 50 mM sodium acetate pH 4.5, 150 mM NaCl containing 2 mM methoxy-capped PEG-

Popp et al.



**Fig. 6.** Synthesis and characterization of cyclic, PEGylated IFN $\alpha$ 2. (A) GG-IFN $\alpha$ -LPETAAH<sub>5c</sub> was used to make backbone-cyclized IFN $\alpha$ 2 with the amino oxy moiety stitched between the site of closure. This material was then used to generate cyclic, PEGylated IFN $\alpha$ 2 by aniline catalyzed oxime ligation. Purified proteins were resolved by 12.5% SDS-PAGE and stained with Coomassie. (B) MS/MS identification of a peptide generated by AspN digestion containing the C terminus of IFN $\alpha$ , followed by the alanine-based probe joined to the N terminus of IFN $\alpha$  by SrtA<sub>Staph</sub>. Note that the amino oxy group was likely lost during MS/MS analysis and ESI-MS reconstructions of unfragmented protein revealed a mass consistent with installation of the amino oxy group (Fig. S3). (C) *In vitro* bioactivity of linear and cyclic IFN $\alpha$  conjugates in a Daudi cell proliferation inhibition assay. Dose-response curves for Daudi cell proliferation inhibition were measured by MTT assay, in triplicate, with the standard deviation displayed. (D) Resistance of circular-AOAA and circular-PEGylated IFN $\alpha$  to thermal denaturation. Change in  $T_m$ 's for circular-AOAA and circular-PEGylated IFN $\alpha$  conjugates, relative to their precursor protein, were measured by thermal shift assay ( $n = 4$ ). (E) Circular cytokines remain biologically active after boiling. The indicated variants were diluted to 1 mg/mL in RPMI/10% serum, boiled for 4 min, allowed to cool at room temperature for 16 h, and assayed for inhibition of Daudi cell proliferation. (F) Change in serum IFN $\alpha$  levels following tail vein injection of circular, site-specifically PEGylated IFN $\alpha$  as well as its linear precursor protein. 10  $\mu$ g of each protein was injected and, at the indicated time points, blood was withdrawn and cytokine concentration in serum was determined by ELISA assay. Data are means from three mice in each group, with the standard deviation displayed.

propionaldehyde, and 100 mM aniline (JT Baker) and incubated at 30  $^{\circ}$ C for 3 h without agitation. This reaction mixture was again purified by cation exchange chromatography, concentrated, and protein concentration was determined by Bradford assay.

**Thermal Denaturation Assays.** Thermal shift assays were performed in a Roche 480 lightcycler and fluorescence was measured with Excitation = 533 nm, Emission = 610 nm. For GCSF-3, 1  $\mu$ g of each variant was mixed with 1  $\mu$ L of 100 $\times$  Sypro orange in 25  $\mu$ L total volume of 50 mM acetate pH 4.0, 150 mM NaCl.  $n = 4$  for all samples with buffer as a blank control and lysozyme as a positive control. For IFN $\alpha$ 2, 1  $\mu$ g of each variant was mixed with

PNAS | February 22, 2011 | vol. 108 | no. 8 | 3173



1  $\mu$ L of 100 $\times$  Sypro orange (Invitrogen) in 25  $\mu$ L total volume of 50 mM MES pH 5.0, 150 mM NaCl.  $n = 4$  for all samples with buffer as a blank control and lysozyme as a positive control. Samples were heated from ambient temperature to 95  $^{\circ}$ C at 0.01  $^{\circ}$ C/s with 6 acquisitions/ $^{\circ}$ C. The included Roche software was used to calculate the negative first derivative of the temperature change as a function of time and the minima were used as the  $T_m$ . Because of the impurities in the cyclic EPO preparations,  $T_m$  was not determined.

**Cell-Based Bioactivity Assays.** Daudi cell proliferation inhibition assays to measure the activity of IFN $\alpha$ 2 conjugates were performed as described (8) and compared to commercial IFN $\alpha$ 2 (PBL Interferon Source) with low passage Daudi cells (ATCC), cultured in RPMI medium 1640/10% heat inactivated fetal calf serum (FCS)/50 units/mL penicillin, 50  $\mu$ g/mL streptomycin sulfate. Cell proliferation was measured by MTT assay according to manufacturer's directions (ATCC). For the denaturation assay, samples diluted to 1  $\mu$ g/mL in RPMI medium 1640/10% FCS were boiled for 4 min, allowed to cool at room temperature for 16 h, and assayed for Daudi cell proliferation inhibition. NFS-60 cell proliferation assays to measure G-CSF conjugate activity were performed as follows. NFS-60 cells (a kind gift from James Ihle, St. Jude Children's Research Hospital) were cultured in RPMI medium 1640/10% FCS/50 units/mL penicillin, 50  $\mu$ g/mL streptomycin sulfate supplemented with murine IL-3 (20 U/mL, R&D Systems). Cells were washed extensively in complete RPMI medium 1640 medium lacking IL-3 and resuspended at  $1 \times 10^5$  cells/mL in complete RPMI medium 1640 lacking IL-3. Titrated G-CSF conjugates (50  $\mu$ L) in complete RPMI medium 1640 were aliquoted into a flat bottom 96 well plate and 50  $\mu$ L of cells ( $0.5 \times 10^4$  cells/well) were added to each well. Cells were cultured for 3 d and an MTT assay (ATCC) was performed according to manufacturer's directions. Each plate contained a titration of commercial G-CSF3 (PeproTech) and readings were blanked against wells containing only NFS-60 cells. The activity of each conjugate was measured a minimum of  $n = 3$  times. BaF3 cells stably expressing the EPO receptor (BaF3 EPO-R cells, a kind gift from Harvey Lodish, Whitehead

Institute) were used to measure EPO activity. BaF3 EPO-R cells were cultured as described for NFS-60 cells and assays were performed essentially as described for G-CSF, except EPO circularization reactions incubated with or without sortase for 16 h at 37  $^{\circ}$ C were used. The total protein concentration of the EPO input material (measured by the Bradford method) was used and was not adjusted for impurities. Commercial preparations of EPO (eBioscience) were used as standards and readings were blanked against BaF3 EPO-R cells cultured with no cytokines. Activities were measured a minimum of  $n = 3$  times.

**Circulatory Half-Life Assays.** For circulatory half-life assays, mice were injected in the tail vein with each protein (10  $\mu$ g per mouse for IFN $\alpha$ 2 conjugates, 5  $\mu$ g per mouse for G-CSF conjugates) and subjected to retroorbital eye bleed at the indicated time points. Blood was harvested, centrifuged at 1,957 $\times$ g in a tabletop centrifuge, and serum was collected and snap-frozen. Elisa assays to measure the quantity of cytokine in serum samples were performed according to the manufacturer's directions (IFN $\alpha$  from PBL Interferon Source, G-CSF from Invitrogen). Each conjugate was injected into  $n = 3$  mice. Data was fit to a two phase exponential decay in GraphPad Prism.

**Granulocyte Proliferation Assay.** Peripheral blood was collected retroorbitally into EDTA collection tubes. Red blood cells were lysed in hypotonic lysis buffer and the remaining peripheral blood mononuclear cells were stained with anti-myeloid differentiation antigen-1 (Gr-1) antibody labeled with phycoerythrin (PE) (BD Pharmingen), anti-cluster of differentiation 11b protein antibody labeled with FITC (anti-CD11b-FITC) (BD Pharmingen), and 7-Aminoactinomycin D (7-AAD) (ViaProbe, BD Pharmingen). Cells were analyzed using a FACS Caliber flow cytometer (BD).

**ACKNOWLEDGMENTS.** We thank Drs. Nick Yoder, John Antos, and Annemarie van der Veen, for helpful discussions, and Drs. Harvey Lodish and James Ihle for reagents. This work was supported by grants from the National Institutes of Health (NIH).

- Wang X, Lupardus P, Laporte SL, Garcia KC (2009) Structural biology of shared cytokine receptors. *Annu Rev Immunol* 27:29–60.
- Leader B, Baca QJ, Golan DE (2008) Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* 7:21–39.
- Hershfield MS, et al. (1991) Use of site-directed mutagenesis to enhance the epitope-shielding effect of covalent modification of proteins with polyethylene glycol. *Proc Natl Acad Sci USA* 88:7185–7189.
- Jevsevar S, Kunsteil M, Porekar VG (2010) PEGylation of therapeutic proteins. *Biotechnol Surv J* 5:113–128.
- Veronese FM (2001) Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials* 22:405–417.
- Grace MJ, et al. (2005) Site of PEGylation and polyethylene glycol molecule size attenuate interferon- $\alpha$  antiviral and antiproliferative activities through the JAK/STAT signaling pathway. *J Biol Chem* 280:6327–6336.
- Doherty DH, et al. (2005) Site-specific PEGylation of engineered cysteine analogues of recombinant human granulocyte-macrophage colony-stimulating factor. *Bioconjugate Chem* 16:1291–1298.
- Rosendahl MS, et al. (2005) A long-acting, highly potent interferon  $\alpha$ -2 conjugate created using site-specific PEGylation. *Bioconjugate Chem* 16:200–207.
- Marraffini LA, Dedit AC, Schneewind D (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* 70:192–221.
- Ton-That H, Liu G, Mazmanian SK, Fahl KF, Schneewind D (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci USA* 96:12424–12429.
- Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL (2007) Sortagging: a versatile method for protein labeling. *Nat Chem Biol* 3:707–708.
- Tanaka T, Yamamoto T, Tsukiji S, Nagamune T (2008) Site-specific protein modification on living cells catalyzed by Sortase. *ChemBiochem* 9:802–807.
- Race PR, et al. (2009) Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J Biol Chem* 284:6924–6933.
- Antos JM, et al. (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* 131:10800–10801.
- Antos JM, et al. (2009) A straight path to circular proteins. *J Biol Chem* 284:16028–16036.
- Bailon P, et al. (2001) Rational design of a potent, long-lasting form of interferon: a 40 kDa branched polyethylene glycol-conjugated interferon  $\alpha$ -2a for the treatment of hepatitis B. *Bioconjugate Chem* 12:195–202.
- Camarero JA, et al. (2001) Rescuing a destabilized protein fold through backbone cyclization. *J Mol Biol* 308:1045–1062.
- Lo MC, et al. (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* 332:153–159.
- Hara K, et al. (1988) Bipotential murine hemopoietic cell line (NFS-60) that is responsive to IL-3, GM-CSF, G-CSF, and erythropoietin. *Exp Hematol* 16:256–261.
- Zhang YL, et al. (2009) Symmetric signaling by an asymmetric 1 erythropoietin: 2 erythropoietin receptor complex. *Mol Cell* 33:266–274.
- Dirksen A, Hackeng TM, Dawson PE (2006) Nucleophilic catalysis of oxime ligation. *Angew Chem Int Ed Engl* 45:7581–7584.
- Kochendoerfer GG, et al. (2003) Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* 299:884–887.
- Shao H, et al. (2005) Site-specific polymer attachment to a CCL-5 (RANTES) analogue by oxime exchange. *J Am Chem Soc* 127:1350–1351.
- Yamamoto T, 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* 9:1022–1024.
- Andersen AS, et al. (2010) Backbone cyclic insulin. *J Pept Sci* 16:473–479.
- Iwai H, Pluckthun A (1999) Circular beta-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett* 459:166–172.
- Clark RJ, et al. (2005) Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. *Proc Natl Acad Sci USA* 102:13767–13772.
- Clark RJ, et al. (2010) The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew Chem Int Ed Engl* 49:6545–6548.
- Trabi M, Craik DJ (2002) Circular proteins—no end in sight. *Trends Biochem Sci* 27:132–138.
- Craik DJ (2006) Chemistry. Seamless proteins tie up their loose ends. *Science* 311:1563–1564.
- Trolier MD, Newkirk MM, King LE, Fraker PJ (2008) Natural glucocorticoids induce expansion of all developmental stages of murine bone marrow granulocytes without inhibiting function. *Proc Natl Acad Sci USA* 105:2028–2033.

**Supplemental Online Material at:**

**[www.pnas.org/content/108/8/3169/suppl/DCSupplemental](http://www.pnas.org/content/108/8/3169/suppl/DCSupplemental)**

**Chapter 7: A system for monitoring influenza  
glycoproteins in living infected cells**

**A system for monitoring influenza glycoproteins in living infected cells**

(Unpublished, Popp M.W., Karssemeijer R., and Ploegh H.L.)

**Abstract**

Influenza glycoproteins control steps during both the entry and the egress of virions and represent the major targets of the humoral immune system. Systems for monitoring the events that contribute to budding of virions from infected cells are lacking. Here we describe a system for monitoring both the hemagglutinin and neuraminidase proteins in living, infected cells during egress of progeny virions.

**Introduction**

Influenza virus is an enveloped, segmented, negative stranded RNA virus encoding 11 proteins, two of which are prominently displayed on the surface of the virion. Flu viruses use the hemagglutinin and neuraminidase glycoproteins for interaction with the host cell membrane<sup>1-2</sup>. Hemagglutinin, a type I transmembrane protein, is synthesized as a single polypeptide, HA<sub>0</sub>, and is cleaved in a 19 amino acid loop by host proteases bearing trypsin-like activity. The HA<sub>0</sub> polypeptide is assembled into non-covalent trimers during biosynthesis, which are then incorporated into nascent virions. Hemagglutinin binds to sialoglycoconjugates on the surface of target cells and the products of the host cell cleavage event, HA<sub>1</sub> and HA<sub>2</sub>, remain disulfide bonded to each other. Following internalization of the virion, hemagglutinin undergoes a dramatic pH dependent conformational change to fully expose HA<sub>2</sub>, leading to fusion between the host and viral membrane and viral genome release<sup>3</sup>.



Chapter 7: A system for monitoring influenza glycoproteins in living infected cells

In contrast, neuraminidase, a type II membrane protein, is involved in cleavage of sialic acid residues during the release of progeny virions from infected cells. It is assembled into disulfide bonded dimers, which are incorporated into virions as non-covalent tetramers<sup>2</sup>.

Entry of viruses is amenable to experimental analysis through the use of chemically or genetically modified virions. Lipophilic dyes, for example, have been used to study the early steps of infection by influenza<sup>4-7</sup>. However, the dearth of labeling methods that can be applied directly to the major viral glycoproteins while preserving their activity has made it impossible to visualize their behavior in living cells and consequently the budding of influenza has not been observed. Live cell imaging studies of other viruses such as vaccinia have used GFP fusions<sup>4, 8-10</sup>, however the hemagglutinin cytoplasmic tail is particularly intolerant of extensions<sup>1, 11-12</sup>. Fusion of GFP to the N-terminus of neuraminidase, a type II membrane protein is likely will impede co-translational insertion into the endoplasmic reticulum.

Due to these limitations, flu budding has been studied mainly by electron microscopy, yielding static pictures of a dynamic process<sup>11, 13-14</sup>. Biochemical assays measuring viral titers and the amount of radiolabeled flu proteins secreted into culture medium<sup>11, 14</sup> have also been used, but by their very nature do not provide resolution at the single particle level. Thus visualization of flu virus budding in real time is an important goal that has been beyond reach due to a lack of tools with which to address the question.

Visualization of the budding process demands a method for discriminating viral proteins

from the many host cell proteins inserted into the very membrane from which viral budding occurs.

We have developed a site-specific protein labeling method that exploits the transpeptidase activity of sortase enzymes to incorporate non-genetically encoded entities into proteins<sup>15</sup>. The sortase enzyme recognizes a five amino acid tag (LPXTG in the case of *Staphylococcus aureus* sortase A), cleaving the recognition sequence between the threonine and glycine with a key cysteine residue and forming an acyl enzyme intermediate. This intermediate is resolved by nucleophilic attack by the N-terminus of a modified oligo-glycine peptide, added exogenously into the labeling mix.

A major advance in influenza biology was the generation of infectious flu particles solely from plasmid DNA<sup>16-18</sup>. These reverse genetics systems have allowed the creation of influenza strains with defined genetic lesions in them, which was not previously possible by reassortment and selection schemes. Using this methodology, we have generated recombinant flu particles that replicate with identical bulk kinetics to their parental strain but contain the requisite LPETG *S. aureus* SrtA cleavage site engineered into either neuraminidase or hemagglutinin. Infection of cells with these recombinant virions leads to production of the flu glycoproteins in functional, taggable forms that can be modified on the surface of the infected cells and are incorporated into mature virions (**Figure 7.1**).

## **Results**

We have employed the 12 plasmid reverse genetics system described by Palese et al<sup>16</sup> to create strains of the A/WSN/33 neurotropic virus that carries sortase cleavage sites on

either the C-terminus of neuraminidase or in the hemagglutinin (HA) protein.

We generated recombinant virions bearing the LPETG cleavage site followed by a hemagglutinin (HA) epitope tag (the A/WSN/33 strain lacks this particular epitope) at the C-terminus of neuraminidase (**Figure 7.5**). This strain (hereafter referred to as NA-Srt) is propagated efficiently in cell culture. Virions pelleted directly from tissue culture supernatant are excellent substrates for SrtA-mediated labeling of neuraminidase (**Figure 7.2**)-sortase compatible probes bearing a biotin (**Figure 7.2a**), tetramethylrhodamine (**Figure 7.2a**) and AlexaFluor647 (**Figure 7.2b**) moieties can all be incorporated, with concomitant loss in the HA epitope downstream from the cleavage site (**Figure 7.2a**). We confirmed that the addition of the sortase cleavage site to the C-terminus of neuraminidase does not affect bulk replication of the modified virus by a multistep replication assay. The modified NA-Srt virus replicates with nearly identical bulk kinetics as the parental WSN strain (**Figure 7.3a**). Plaque morphology of the NA-Srt virus is also similar to wild-type (**Figure 7.3b**). The biosynthesis of neuraminidase is also unaffected by the tag-virions pelleted from tissue culture supernatant and further purified through a sucrose gradient show only quantitatively disulfide bonded dimers of neuraminidase are incorporated into virus (**Figure 7.3c**).

Infected Madin-Darby Canine Kidney (MDCK) cell surfaces can also be labeled selectively, as demonstrated by incorporation of a biotin probe into surface-displayed neuraminidase (**Figure 7.4**). These data establish that the modified virus is infectious, and that neuraminidase can be labeled on the surface of both virions and infected cells.

To identify a position in HA amenable to sortase cleavage, we generated three different constructs (**Figure 7.5**), two with the sortase cleavage site, LPETGG, either substituted or inserted upstream of the HA<sub>1</sub>-HA<sub>2</sub> trypsin cleavage site and also a version where the more conservative LPSTG substitution was made. We failed to generate recombinant virus with the LPSTG substitution, but succeeded in rescuing virions from the two LPETGG tagged strains. Hemagglutinin incorporated into virions can be labeled by sortase when virions are pelleted directly from supernatant derived from the mixture of cells used to rescue virus (**Figure 7.6**), however when this supernatant was used to re-infect MDCK cells, only cells infected with the recombinant virus bearing HA with the LPETGG insertion (hereafter termed HA-Srt virus) showed cell-surface labeling of hemagglutinin using sortase and a biotinylated probe (**Figure 7.7**). We characterized the replication of the HA-Srt virus by a multi-step replication assay using hemagglutination of chicken erythrocytes as a proxy for viral load in the supernatant. Like NA-Srt, the HA-Srt virus replicated with nearly identical kinetics to the parental WSN strain (**Figure 7.8**).

To confirm that the influenza proteins produced by cells infected with either the NA-Srt or HA-Srt viruses are glycosylated, we subjected the surfaces of infected cells to sortase-labeling with a biotinylated probe, a procedure that will label only the surface displayed protein poised for incorporation into virions. Digestion of cell lysates with either Endoglycosidase H or PNGaseF revealed that surface exposed biotinylated HA-Srt protein was largely EndoH resistant, indicating successful traversal of the secretory pathway and maturation of N-linked oligosaccharides (**Figure 7.9**). The surface

displayed neuraminidase produced during infection is partially EndoH sensitive, however, an observation consistent with literature reports (**Figure 7.9**).

We further characterized the NA-Srt virus by infecting mice with a sub-lethal dose of the recombinant and wild-type parental strains. Both influenza strains caused weight loss in mice, followed by recovery (**Figure 7.10**), indicating that the NA-Srt strain maintains the ability to infect living animals despite addition of the sortase recognition element to neuraminidase.

MDCK cell surfaces infected with either the NA-Srt or HA-Srt viruses can be labeled (**Figure 7.4** and **Figure 7.7**). Neither the sortase enzyme nor the peptide probes used are cell permeable, thus confining labeling to the cell surface pool of protein. For the NA-Srt strain, this subset of protein is with slightly delayed kinetics relative to biosynthesis of the protein after infection (**Figure 7.4**, compare top and middle panels) and labeling is absolutely selective; no endogenous proteins are labeled. Given these observations, we examined whether labeled neuraminidase in infected cells could be observed by microscopy (**Figure 7.1**). Unpolarized MDCK cells seeded on glass coverslips were infected with NA-Srt virus, labeled with sortase and a glycine based probe bearing an AlexaFluor647 dye, and imaged by confocal microscopy. A series of images in the Z plane shows the expected cell surface disposition of the neuraminidase (**Figure 7.11**) as evidenced by the rim-like staining of cells. At the top of the Z-series, we observe punctate staining (**Figure 7.11, right most panels**), a pattern that we infer to be nascent influenza budding sites.

## **Discussion**

We have devised a method for observing the behavior of influenza glycoproteins in cells infected with a fully functional virus. Our method leverages the sortase labeling technology<sup>15, 19</sup> as well as influenza viruses engineered to carry short sortase recognition sequences in their glycoproteins. These (minimally) engineered viruses show identical behavior to the parental strain with respect to infectivity, replication kinetics, and protein biosynthesis and thus observation of the sortase labeled glycoproteins reflects the behavior of their wild-type counterparts. By combining SrtA-labeling methodology with live-cell imaging, we will monitor the kinetics of influenza budding and release in future studies. We anticipate that this system will yield a robust method to visualize the kinetics of particle formation and, in combination with perturbations of host cells, will reveal host proteins that contribute to the process of influenza virion biogenesis.

## **Methods**

**Protein production and peptide probe synthesis.** Sortase was produced as described<sup>15</sup>. Peptide probes were produced as described<sup>15</sup> (also see **Appendix A**).

**Generation of recombinant viruses.** Mutant viruses were generated by reverse genetics using plasmids described<sup>16</sup>. The hemagglutinin and neuraminidase plasmids were modified by standard molecular biology techniques to carry the sortase cleavage site. All viruses, including the wild-type WSN virus used were rescued as described<sup>16</sup>.

Chapter 7: A system for monitoring influenza glycoproteins in living infected cells

**Viral Assays.** Viral titer was assessed by plaque assay on MDCK cells as described<sup>14</sup>. For multi-step replication assays, MDBK cells were infected at an MOI of 0.001 and incubated for the indicated times in viral growth medium (VGM, DME with 0.3% BSA) supplemented with 0.5 µg/ml TPCK-treated trypsin (NA-Srt) or 1 µg/ml TPCK treated trypsin (HA-Srt). For NA-Srt multi-step replication, 100 µl of media was plaqued at the indicated time points on MDCK cells grown with 0.5 µg/ml TPCK-trypsin. For HA-Srt, hemagglutination using chicken erythrocytes (Lampire labs) was done at the indicated time points, as described<sup>16</sup>.

**Viral Particle Purification.** Viral particles from tissue culture supernatant were concentrated by pelleting through a 20% sucrose cushion (Sigma) at 25000 rpm in an SW-28 rotor for 120 minutes. Where indicated, virus was further purified through a continuous 15%-60% sucrose gradient, centrifuged at 30,200 rpm in an SW-40.1 rotor for 3 hours.

**Sortase labeling of virions.** For biotin and TAMRA labeling, virions were pelleted from tissue culture supernatant without a sucrose cushion. The pellet was resuspended in 1X sortase buffer<sup>15</sup> and labeled with 150 µM sortase A/ 5 mM probe for 1 hours at 37°C. For Alexa647 probe labeling, sucrose gradient purified virions were mixed with 200 µM sortase A and 500 µM probe at 37°C for 2 hours.

**Sortase labeling of live cell surfaces for immunoblot.** MDCK cells were plated in a 24 well dish at 70% confluency the night before the experiment. Cells were infected at an

*Chapter 7: A system for monitoring influenza glycoproteins in living infected cells*

MOI of 1 and 0,1,2,3,4,5,6 or 7 hours post infection, cells were incubated with 100mM sortase A and 100mM G<sub>5</sub>K-biotin probe<sup>15</sup> in VGM for 30 minutes at 37°C. Cells were washed extensively in PBS, collected, and lysed in 1% SDS with protease inhibitor cocktail (Roche). A BCA assay was performed (Pierce) and 20 µg of lysate was loaded for western blotting.

**Mouse infections.** Mice (n=4 in each group) were inoculated intranasally with 40000 pfu of the indicated virus and body weight was monitored at the indicated intervals.

**Cell surface labeling and glycosidase digestion.** MDCK cells were infected at an MOI of 0.4 overnight and labeled for 1 hour at 37°C with 100 mM sortase and 500 mM biotin probe. Cells were then lysed in glycoprotein denaturing buffer (New England Biolabs) and total protein in lysates were quantitated by BCA assay (Pierce). Five micrograms of cell lysate was digested with either PNGase F or EndoH according to manufacturer's directions (New England Biolabs), resolved by 12.5% SDS-PAGE, transferred to nitrocellulose, and used for western blotting with the indicated antibodies.

**Imaging.** MDCK cells were seeded onto 18 mm circular glass coverslips (VWR International) and infected with NA-Srt at an MOI of 0.5. At 7.5 hours post-infection, coverslips were inverted and placed onto a drop of labeling mix (200 µM SrtA and 100 µM Alexa647 probe in VGM) for 10 minutes at 37°C. Coverslips were then washed 5 times in PBS, and fixed by inversion onto a drop of 4 % paraformaldehyde in PBS for 10 minutes at room temperature on parafilm. Coverslips were mounted with Fluoromount G



Chapter 7: A system for monitoring influenza glycoproteins in living infected cells

(Southern Biotech) and analyzed using a Nikon spinning disk confocal microscope with Metamorph software.

## References

1. Nayak, D.P., Hui, E.K. & Barman, S. Assembly and budding of influenza virus. *Virus Res* 106, 147-165 (2004).
2. Steinhauer, D.A. & Skehel, J.J. Genetics of influenza viruses. *Annu Rev Genet* 36, 305-332 (2002).
3. Harrison, S.C. Viral membrane fusion. *Nat Struct Mol Biol* 15, 690-698 (2008).
4. Brandenburg, B. & Zhuang, X. Virus trafficking - learning from single-virus tracking. *Nat Rev Microbiol* 5, 197-208 (2007).
5. Lakadamyali, M., Rust, M.J., Babcock, H.P. & Zhuang, X. Visualizing infection of individual influenza viruses. *Proc Natl Acad Sci U S A* 100, 9280-9285 (2003).
6. Lakadamyali, M., Rust, M.J. & Zhuang, X. Endocytosis of influenza viruses. *Microbes Infect* 6, 929-936 (2004).
7. Rust, M.J., Lakadamyali, M., Zhang, F. & Zhuang, X. Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat Struct Mol Biol* 11, 567-573 (2004).
8. Ward, B.M. Pox, dyes, and videotape: making movies of GFP-labeled vaccinia virus. *Methods Mol Biol* 269, 205-218 (2004).
9. Ward, B.M. Visualization and characterization of the intracellular movement of vaccinia virus intracellular mature virions. *J Virol* 79, 4755-4763 (2005).
10. Jouvenet, N., Bieniasz, P.D. & Simon, S.M. Imaging the biogenesis of individual HIV-1 virions in live cells. *Nature* 454, 236-240 (2008).
11. Barman, S. et al. Role of transmembrane domain and cytoplasmic tail amino acid sequences of influenza A virus neuraminidase in raft association and virus budding. *J Virol* 78, 5258-5269 (2004).
12. Ohuchi, M., Fischer, C., Ohuchi, R., Herwig, A. & Klenk, H.D. Elongation of the cytoplasmic tail interferes with the fusion activity of influenza virus hemagglutinin. *J Virol* 72, 3554-3559 (1998).
13. Wang, X., Hinson, E.R. & Cresswell, P. The interferon-inducible protein viperin inhibits influenza virus release by perturbing lipid rafts. *Cell Host Microbe* 2, 96-105 (2007).
14. Barman, S. & Nayak, D.P. Lipid raft disruption by cholesterol depletion enhances influenza A virus budding from MDCK cells. *J Virol* 81, 12169-12178 (2007).
15. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* 3, 707-708 (2007).
16. Fodor, E. et al. Rescue of influenza A virus from recombinant DNA. *J Virol* 73, 9679-9682 (1999).
17. Neumann, G. & Kawaoka, Y. Reverse genetics systems for the generation of segmented negative-sense RNA viruses entirely from cloned cDNA. *Curr Top Microbiol Immunol* 283, 43-60 (2004).
18. Neumann, G. et al. Generation of influenza A viruses entirely from cloned cDNAs. *Proc Natl Acad Sci U S A* 96, 9345-9350 (1999).
19. Tanaka, T., Yamamoto, T., Tsukiji, S. & Nagamune, T. Site-specific protein modification on living cells catalyzed by Sortase. *ChemBiochem* 9, 802-807 (2008).

## **Figure Legends**

### **Figure 7.1. Scheme for studying influenza glycoprotein behavior in live infected cells**

Observation by live microscopy of influenza infected cells demands a method for the site-specific installation of probes onto flu proteins at the exclusion of host-cell proteins (gray squares). We have generated recombinant influenza viruses that encode, in their genome, versions of neuraminidase (NA, round circles) or hemagglutinin (HA, diamonds) that can be labeled with sortase A (see text for details). Live, infected cells are labeled by inclusion of sortase A and a oligo-glycine based probe in the culture medium.

### **Figure 7.2. Recombinant NA-Srt virions are substrates for sortase A**

(a) NA-Srt virus bearing a sortase cleavage site appended to the C-terminus of neuraminidase, followed by an HA epitope, was pelleted from tissue culture supernatant and incubated with 150  $\mu$ M sortase A and 5 mM of the indicated probe at 37°C for 1 hour, with agitation. Streptavidin-HRP immunoblot (top panel) shows incorporation of a G<sub>5</sub>K(Biotin) probe with concomitant loss of the HA epitope (middle blot). A glycine based tetramethylrhodamine probe can also be incorporated (bottom panel, fluorescence gel).

(b) Highly purified NA-Srt virus is labeled with a glycine based AlexaFluor 647 probe. NA-Srt virions were purified through a 20% sucrose cushion and further banded on a continuous 15-60% sucrose gradient before labeling with sortase (200  $\mu$ M) and an AlexaFluor 647 probe (500  $\mu$ M) for 2 hours at 37°C. The labeling reaction was

fractionated by 12.5% SDS-PAGE and scanned by a Typhoon Imager.

**Figure 7.3. NA-Srt Virus is not attenuated *in vitro***

- (a) Multi-step replication assay. NA-Srt virus and wild-type WSN virus were used to infect MDBK monolayers at an MOI=0.001 and viral supernatant was plaqued on MDCK cells at the indicated times.
- (b) Plaque morphology of WSN and NA-Srt viruses.
- (c) NA-Srt protein incorporated into virions is quantitatively forms disulfide bonded dimers. NA-Srt virus was purified over a sucrose gradient and loaded on a 12.5% SDS-PAGE gel in the presence or absence of reducing agent (2-mercaptoethanol). Anti-HA epitope immunoblot displays the presence of the NA-Srt protein.

**Figure 7.4. Cell surface labeling of NA-Srt infected MDCK cells.**

Neuraminidase on infected MDCK cell surfaces is selectively labeled by SrtA. MDCK cells were infected at MOI = 0.5 and at the indicated times post infection, were incubated for 30 min with 200  $\mu$ M SrtA and 500  $\mu$ M biotinylated nucleophile. Cells were collected, subjected to 12.5% SDS-PAGE, and immunoblots with streptavidin-HRP (top), anti-HA (middle), and anti-p97 (bottom, loading control) were performed.

**Figure 7.5. Protein Sequences**

Protein segments for hemagglutinin and neuraminidase are displayed along with the sequences of the mutant viruses. Sortase cleavage sites are in red, the endogenous trypsin cleavage site on hemagglutinin is in blue and the HA epitope appended to the C-terminus

of the NA-Srt protein is in green.

**Figure 7.6. Rescue and labeling of sortase-compatible hemagglutinin mutant viruses.**

Viruses with the indicated mutations were rescued as described (refs). Supernatant (10 mL) from the cell mixture used to rescue virus was ultracentrifuged and the pellets were incubated with sortase A (200  $\mu$ M) and Biotin probe (1 mM) for 5 hours at 37°C.

Reactions were then denatured and digested with PNGaseF for 1 hour at 37°C to display the denuded, labeled, HA polypeptides and distinguish it from possible NA-Srt virus contamination.

**Figure 7.7. Cell surface labeling of HA-Srt infected MDCK cells.**

Hemagglutinin on infected MDCK cell surfaces is selectively labeled by SrtA. MDCK cells were infected at an MOI = 1 for NA-Srt virus and with supernatant from the cells used to rescue the initiated hemagglutinin mutant viruses. At the indicated times post infection, cells were incubated for 30 min with 200  $\mu$ M SrtA and 500  $\mu$ M biotinylated nucleophile. Cells were collected, subjected to 12.5% SDS-PAGE, and immunoblots with streptavidin-HRP (top), or a mixture of anti-HA, and anti-p97 (loading control) antibodies (bottom) were performed.

**Figure 7.8. HA-Srt virus is not attenuated *in-vitro*.**

Multi-step replication assay. HA-Srt virus and wild-type WSN virus were used to infect MDBK monolayers at an MOI=0.001 and viral supernatant was subjected to

hemagglutination assay with chicken red blood cells at the indicated times.

**Figure 7.9. Hemagglutinin and neuraminidase from HA-Srt and NA-Srt viruses are glycosylated.**

MDCK cells were infected at an MOI of 0.4 overnight and labeled for 1 hour at 37°C with 100 mM sortase and 500 mM biotin probe. Cells were then lysed in glycoprotein denaturing buffer (New England Biolabs) and total protein in lysates were quantitated by BCA assay (Pierce). Five micrograms of cell lysate was digested with either PNGase F or EndoH according to manufacturer's directions (New England Biolabs), resolved by 12.5% SDS-PAGE, transferred to nitrocellulose, and used for western blotting with the indicated antibodies

**Figure 7.10. NA-Srt virus is not attenuated *in-vivo*.**

Mice (n=4 in each group) were inoculated with 40000 pfu of the indicated virus and body weight was monitored at the indicated intervals.

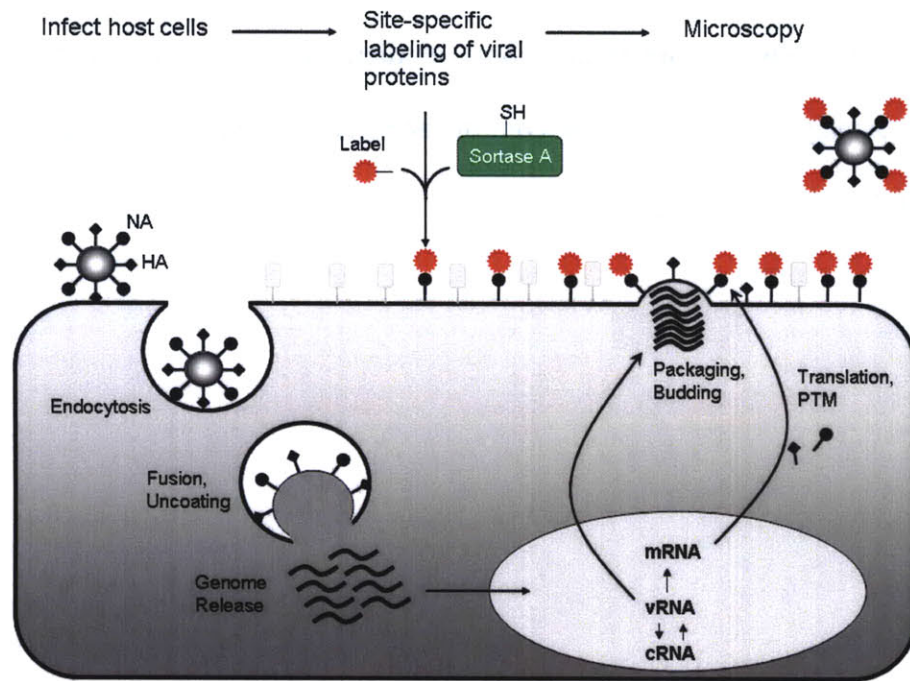
**Figure 7.11. Microscopic observation of labeled neuraminidase on infected cell surfaces.**

MDCK cells were seeded onto coverslips and infected with NA-Srt at an MOI of 0.5. At 7.5 hours post-infection, coverslips were inverted and placed onto a drop of labeling mix (200 μM SrtA and 100 μM Alexa647 probe in VGM) with (top two rows) or without sortase (bottom row) for 10 minutes at 37°C. Coverslips were then washed 5 times in PBS, and fixed by inversion onto a drop of 4 % paraformaldehyde in PBS for 10 minutes

*Chapter 7: A system for monitoring influenza glycoproteins in living infected cells*

at room temperature on parafilm. Coverslips were mounted with Fluoromount and analyzed using a Nikon spinning disk confocal microscope with Metamorph software. Z-stacks are displayed. The no sortase control panels are overexposed to highlight background.

**Figure 7.1**





Chapter 7: A system for monitoring influenza glycoproteins in living infected cells

**Figure 7.2**

a                      WSN                      NA-Ert                      b                      N1-Srt

**Figure 7.3**

**a**

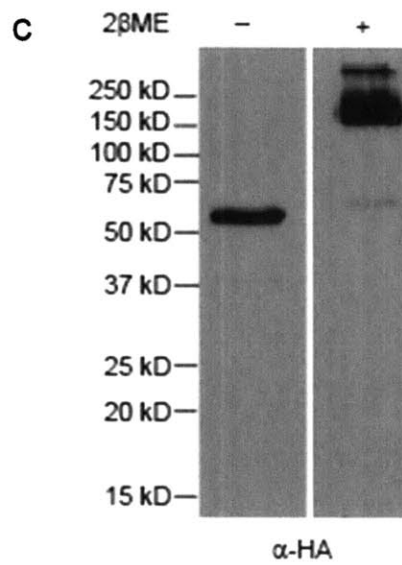
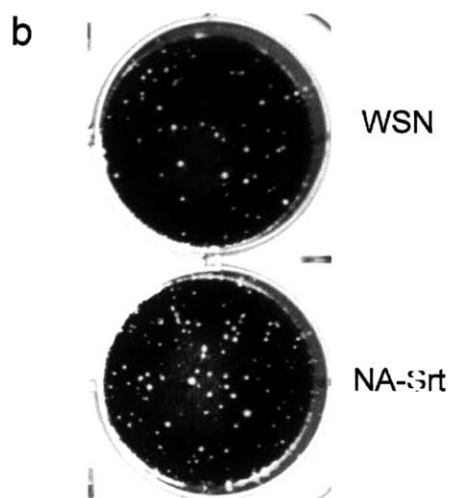
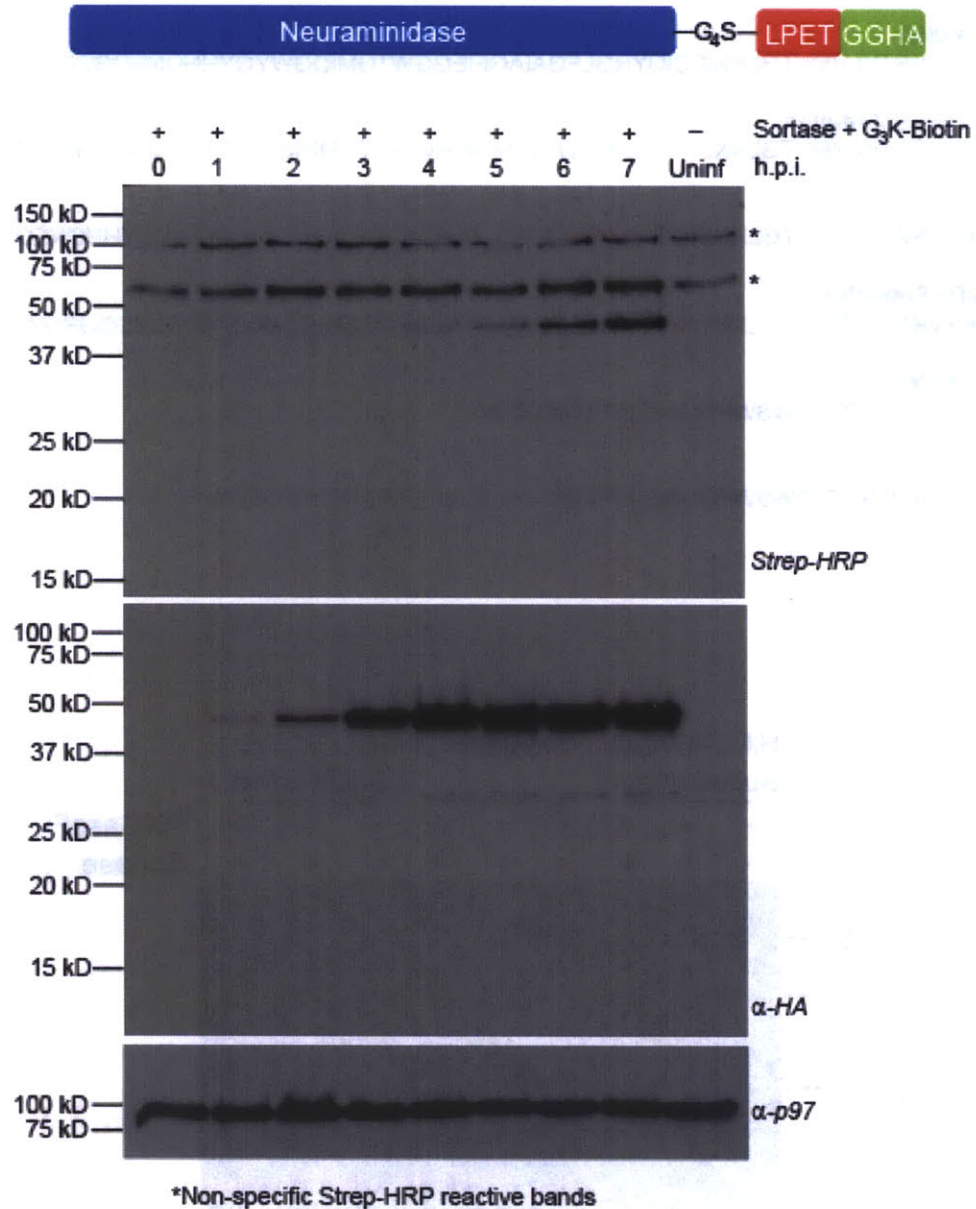


Fig. 7.4



**Fig. 7.5**

**HA WT (WSN)**

319-CPKYVRSTKLRMVTGLRNIPSIQY**RGL**FGAIAGFIEGGWTGMIDGWYGYHHQNEQGSYAA-379

**HA LPETGG Substitution**

319-CPKYVRSTKLRMVTGLRN**LPETGG****RGL**FGAIAGFIEGGWTGMIDGWYGYHHQNEQGSYAA-

**HA-Srt**

319-CPKYVRSTKLRMVTGLRNIPSIQY**LPETGG****RGL**FGAIAGFIEGGWTGMIDGWYGYHHQNEQGSYAA-

**HA LPSTG Substitution**

319-CPKYVRSTKLRMVTGLRN**LPSTG****RGL**FGAIAGFIEGGWTGMIDGWYGYHHQNEQGSYAA-

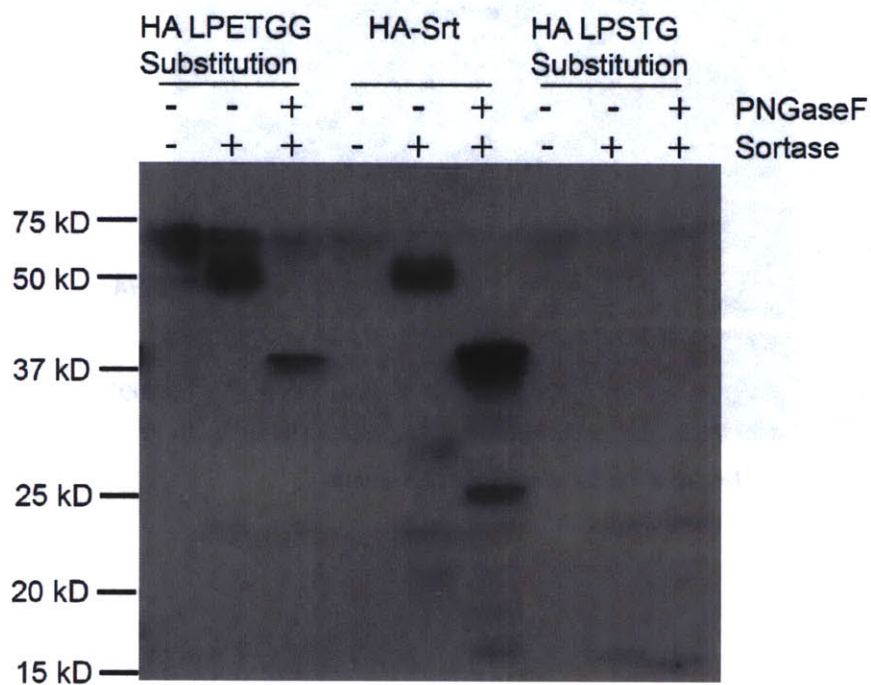
**NA WT (WSN)**

425-SIISFCGVNGDVTVDWSWPDGAELPFTIDK\*-453

**NA-Srt**

425-SIISFCGVNGDVTVDWSWPDGAELPFTIDK**GGGG****SLPETGG****YPYDVPDYA**\*

**Figure 7.6**



**Figure 7.7**

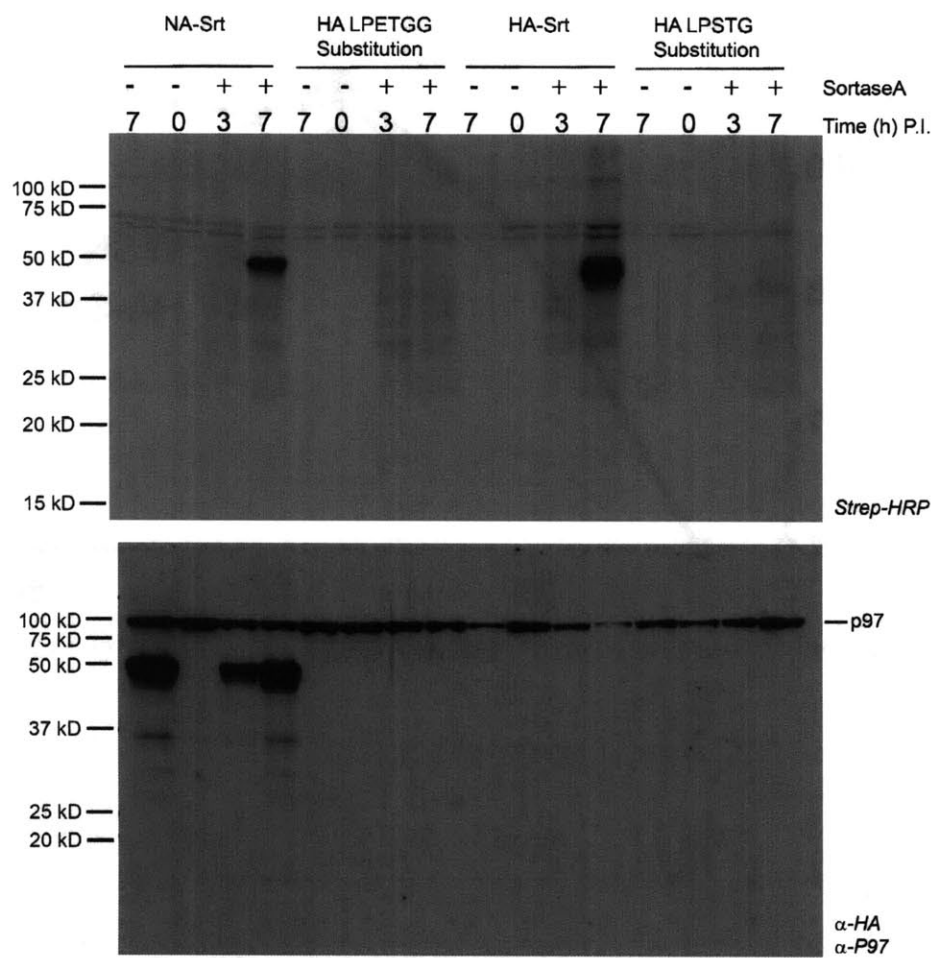


Fig. 7.8

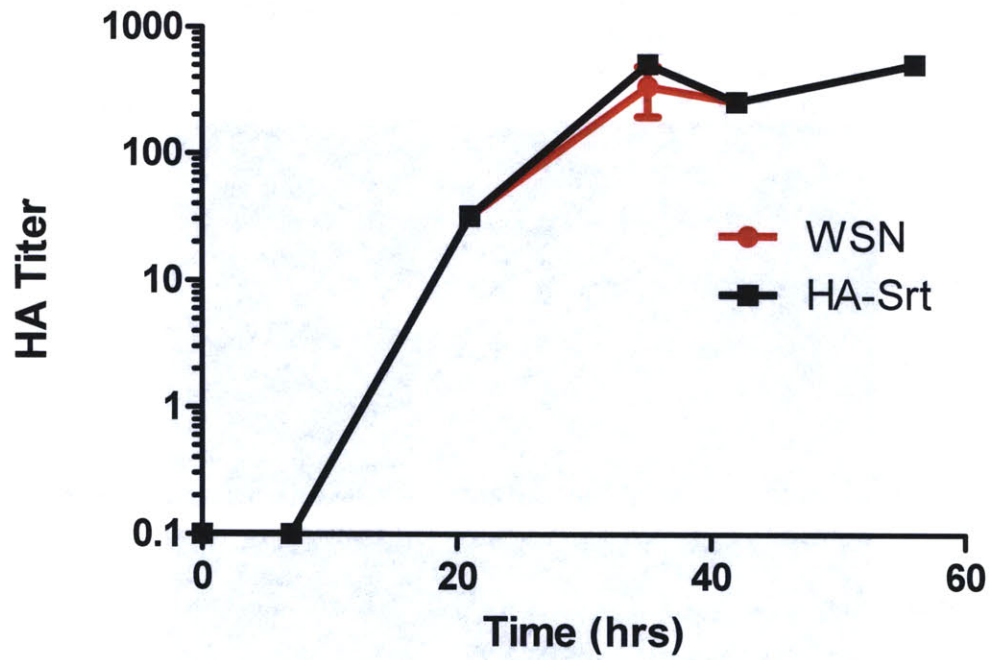


Fig. 7.9

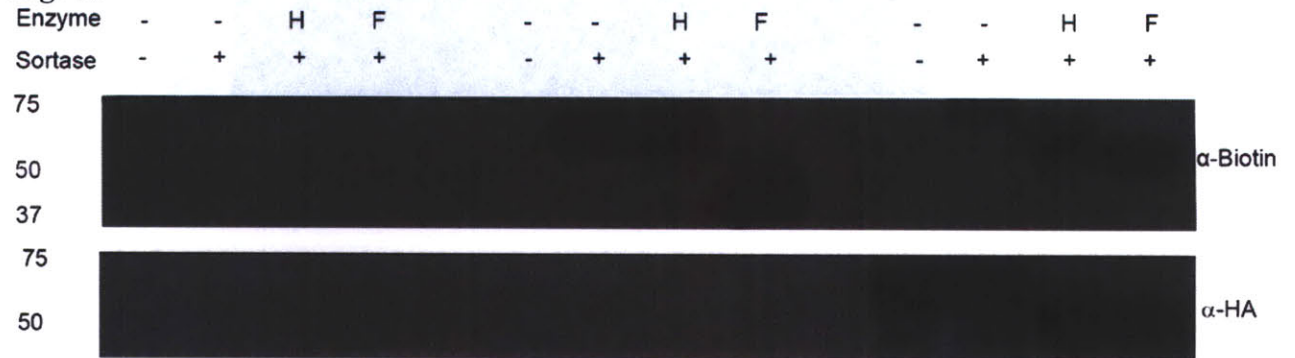


Figure 7.10

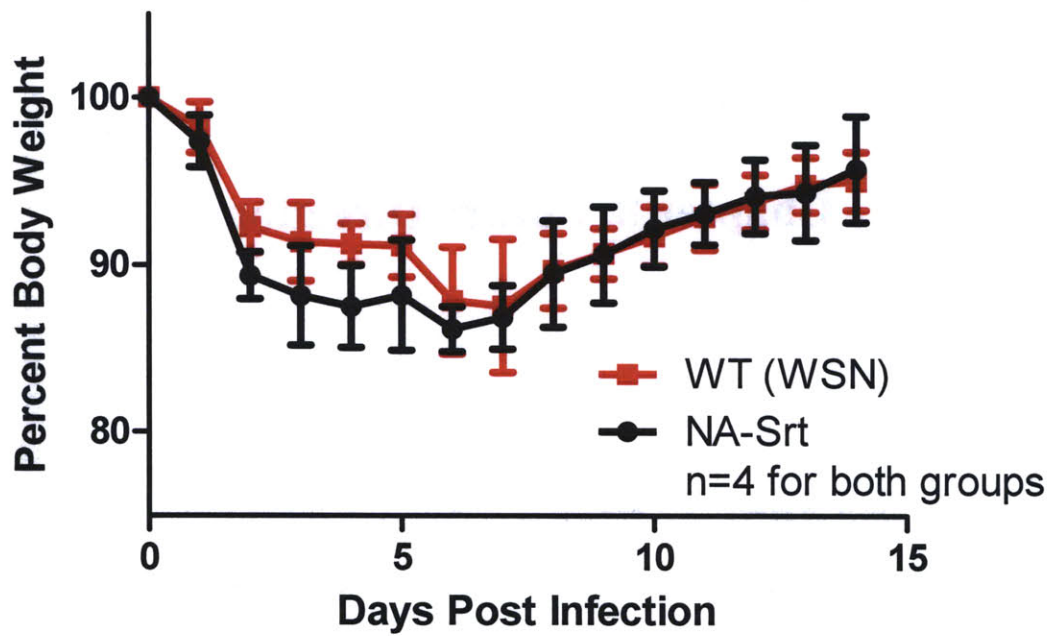
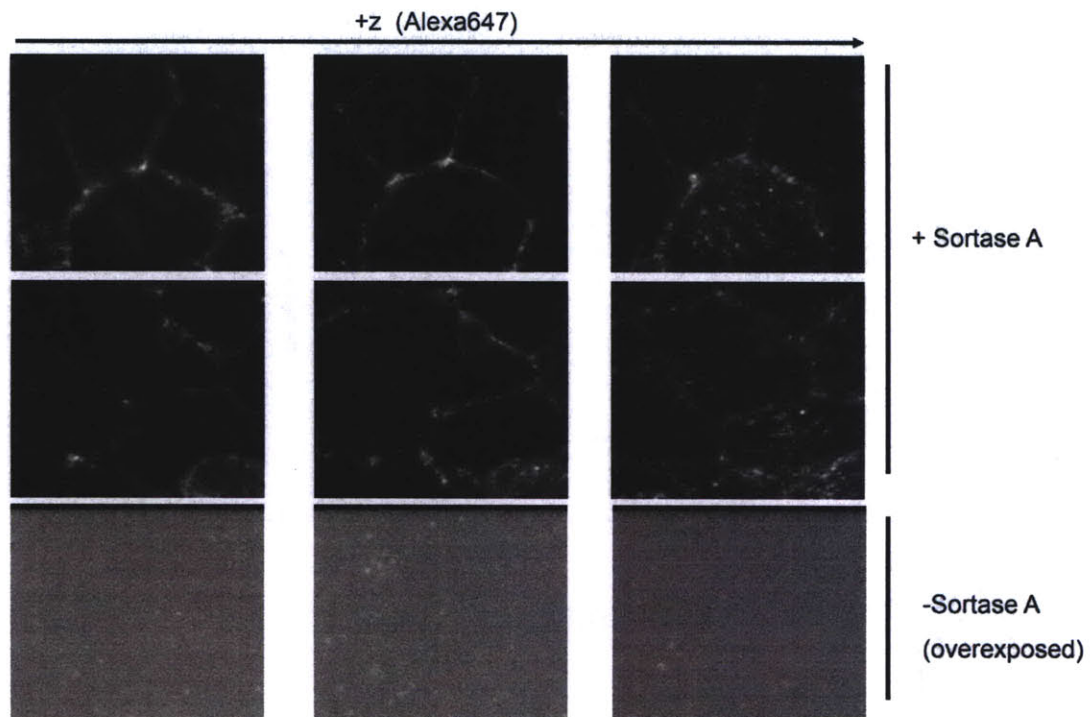


Fig. 7.11



## **Chapter 8: Future Directions**



## Chapter 8: Future Directions

### **Chapter 8: Future Directions**

(Adapted from: Popp M.W., Ploegh H.L. Making and Breaking Peptide Bonds: Protein engineering using sortase, *Angewandte Chemie, Accepted*)

#### **New Technology**

##### **Orthogonal Sortases**

The diversity of the microbial world includes multiple sortase-type enzymes, with several distinct cleavage site preferences and nucleophile specificities.<sup>1</sup> The pilin-building sortases, for example, accept the  $\epsilon$ -amine of a lysine within the YPKN sequence as nucleophile, whereas sortase A from *S. aureus* will accept the N-terminus of glycine extensions. Sortase B (SrtB) from *S. aureus* cleaves the NPQTN sequence,<sup>2-3</sup> while SrtB from *Bacillus anthracis* cleaves the NPKTG sequence<sup>4</sup> and accepts *meso*-diaminopimelic acid as a nucleophile.<sup>5-7</sup> Although the use of distinct sortases for protein labeling is in its infancy, there is clearly great untapped potential in these other sortases if the recognition requirements and reaction conditions can be more clearly specified. With the advent of many directed protein evolution techniques, it may be possible to improve on what nature has provided. In its usual biological context, the housekeeping sortase anchors to the cell wall a set of substrates that are structurally distinct:<sup>8</sup> the one feature required for attack by sortase seems to be the LPXTG pentapeptide. This, coupled with the fact that substrate recognition is primarily a function of the  $\beta$ 6- $\beta$ 7 loop,<sup>9-11</sup> implies that it should be possible to evolve versions of sortase that will attack fully orthogonal peptide sequence motifs. Given the diversity of sortases that recognize different anchor structures,<sup>5-6</sup> it should be possible to evolve enzymes that will accept nucleophiles other than those that occur naturally. By combining sortases of distinct specificity, one can envision orthogonal labeling of different proteins in a one-pot reaction, or -its biological equivalent- inserted into the same cellular environment. The kinetics of the sortase reaction are lackluster at best, most likely

## Chapter 8: Future Directions

owing to the active site equilibrium which favors the inactive, non-ionized cysteine residue. Although this problem is easily overcome by adding large quantities of sortase to the reaction (and thus adding a larger total amount of catalytically competent enzyme), it would still be beneficial to improve the overall kinetics of transpeptidation by *S. aureus* SrtA using directed evolution.

The pilus-building sortases which polymerize subunits by creating iso-peptide bonds remain to be exploited. These enzymes may be of particular interest, since they evolved to establish covalent protein-protein linkages post-translationally, and perhaps this trait could be harnessed to perform similar protein-protein ligations *in vitro*, exploiting exposed lysine residues internal to the protein of interest as the incoming nucleophile. Given that isopeptide bond formation is featured in many biological processes, for example in the ubiquitin system, the ability to make homogeneous, site-specific isopeptide linkages would be tremendously useful.

### **Intracellular Labeling**

An intriguing question is how to overcome the limitations that prevent intracellular labeling by sortase. It is possible to express sortase inside of both living bacterial and eukaryotic cells - the enzyme folds properly and is active (Popp, Strijbis, unpublished observations). The *S. pyogenes* enzyme, lacking the requirement for divalent cations, is active in compartments devoid of calcium, such as the cytosol. Similarly, substrates that bear an exposed LPXTG motif are easily expressed, and depending on the protein, retain their normal localization and/or function. When the nucleophile for transpeptidation is genetically encoded as part of the substrate protein, as in the case of circularization or protein-protein ligation reactions, transpeptidation proceeds readily

## Chapter 8: Future Directions

(Strijbis, Popp, unpublished observations), and yields circular proteins intracellularly in yeast, mammalian cells, and prokaryotes. The limiting step for simple C-terminal labeling is thus delivery of an exogenous nucleophile to the cell interior in quantities that permit biochemical analysis of the transpeptidation product. These limitations might be overcome by attaching one of the many protein transduction sequences derived from the human immunodeficiency virus trans-activator of transcription protein (HIV TAT) or other sources to the labeled nucleophile of interest, although this possibility remains to be reduced to practice and fails to address the fate of any unincorporated cytoplasmically delivered nucleophile. Alternatively, it might be possible to transiently mask the alpha-amine group and other groups that preclude delivery to the cytosol through installation of chemical entities that can be removed by esterases, photochemically, or in the reducing environment of the cytoplasm, making probe delivery irreversible.

### **Applications**

#### **Polarized cell protein transport**

With the tools developed in this thesis, experiments to address many unanswered questions in glycoprotein trafficking and quality control become immediately accessible. Transport of transmembrane glycoproteins within the cellular membranes of polarized cells—the normal context for many host and microbial proteins involved in pathogenesis—is currently studied by crude biochemical assays. For most proteins, it is thought that they are sorted and inserted into their destination membrane during biosynthesis, while other types of proteins are inserted into both membranes and localized to their final destination via transcytosis<sup>12</sup>. The issue of protein sorting and localization is particularly salient for studying influenza infection, where budding usually occurs at the apical domain of polarized epithelial cells. However, to assess this

## Chapter 8: Future Directions

phenomenon requires meticulous biochemical assays-usually a biosynthetic pulse-chase labeling of polarized cells combined with vectorial delivery of non-specific chemical biotinylation reagents to uniformly label all surface proteins on a given surface. Transcytosis is indirectly inferred by the transient accumulation of a protein on a given surface, followed by a steady decline in the face of constant accumulation on the opposite surface. By labeling suitably tagged glycoproteins in polarized cells, it is possible to directly study the behavior of the labeled protein. One need only deliver the sortase labeling mix to the desired surface in order to visualize the fate of such proteins in real time.

### **Defining protein complexes**

Epitope tagged versions of proteins in combination with immunoprecipitation and mass-spectrometry have been used with great success to identify the members of multiprotein complexes. However, this method is limited by the inclusion of detergent, which almost certainly disrupts weak or transient protein-protein interactions. To overcome this significant technical hurdle, sortase can be used to deliver photo-activatable crosslinker probes (see **Chapter 2**) to cell surfaces, semi-intact cells (Popp, unpublished data) or to microsomal preparations containing LPETG tagged bait proteins, in the absence of detergent. In this way, it should be possible to both identify proximity relationships between known interactors as well as identify new proteins in the same complex (**Figure 8.1**). Proximity relationships between known interactors can also be defined by co-expressing a protein C-terminally fused to the sortase cleavage site with a candidate protein containing a free glycine residue at the N-terminus. Delivery of sortase will covalently crosslink these proteins, provided that they are in close enough proximity to one another (see p97 experiments, **Chapter 4**).

## Chapter 8: Future Directions

### **Transient cell-cell interactions**

There is a dearth of tools that report on cell-cell interactions. Currently available split-GFP reporters only detect long-lasting interactions<sup>13</sup> and a very recently developed method for enzymatic biotinylation still relies upon streptavidin linked detection<sup>14</sup>. We envision a tool for reporting on transient cell-cell interactions, where one cell bearing a sortase substrate comes into close proximity to another cell bearing a membrane anchored version of either SrtA<sub>strep</sub> or SrtA<sub>staph</sub>. This would allow for acyl-enzyme formation between the sortase and substrate on opposite cells and when incubated with a functionalized probe (either based on an AA nucleophile or a GG nucleophile for SrtA<sub>strep</sub> and SrtA<sub>staph</sub>, respectively) will ultimately result in labeling of the cell bearing the sortase substrate. In this way an indelible mark for a possibly transient cell-cell interaction is imparted on the cell bearing the sortase substrate, given that the interaction is long-lived enough for the transacylation reaction to occur (**Figure 8.2**).

In its natural context, sortases are embedded in the bacterial membrane through a transmembrane anchor at the N-terminus. We can thus express sortase at the mammalian cell surface by fusing the catalytic domain of sortase to a mammalian type II membrane protein bearing a signal-anchor sequence for ER insertion. This results in insertion into the plasma membrane and exposure of the sortase catalytic domain to the extracellular space (Popp, unpublished data).

### **Bioconjugate Libraries**

A key advantage of the sortase labeling method is the relative ease with which the nucleophilic probes are synthesized. Most probes are simple C-terminal appendages of glycine, made on solid phase, with the reporter either incorporated while the peptide is on resin or in solution by standard chemical coupling techniques. The published record shows that a diverse array of such

## Chapter 8: Future Directions

probes will all likely react as nucleophiles with a given target in the same manner. This makes it possible to create libraries of bioconjugates without the need to re-optimize ligation conditions for each conjugate individually. Cell- and tissue-specific proteinaceous delivery vehicles for small interfering RNA (siRNA) libraries and antibody-drug library conjugates are all within the realm of possibility. An added benefit of the sortase method is the tight control over the stoichiometry of labeling-this should greatly facilitate quantitative studies of the effects of such bioconjugate libraries.

### **Non-natural Protein Topologies**

In addition to the strict control over stoichiometry of ligation products, sortase allows control over the orientation between ligation partners, and this could be useful in creating protein-protein conjugates. By switching the location of the N-terminal glycines on one unit for the sortase motif on the other partner protein, the orientation of the subunits can be reversed. By using cleverly designed probes bearing handles for bioorthogonal reactions (for example azide-alkyne, azide-strained cyclooctyne, or aldehyde-amino-oxy groups), it should be possible to construct protein pairs that are covalently linked in non-natural topologies (N-N or C-C). Alternatively, two-step transacylation procedures using the semi-orthogonal SrtA<sub>Staph</sub> and SrtA<sub>Strep</sub> enzymes along with the appropriate peptide probes in N-N or C-C linkage can be employed. The sortase method thus provides access to protein linkages not found in nature and one area where this can be applied is in making engineered ligands for studying cell signaling. It would be possible to create chimeric, folded protein ligands that constrain binding and heterodimerization of their cognate receptors. It is clear that the ease of use and flexibility of the sortase labeling method will provide exciting new avenues of research.

## Chapter 8: Future Directions

### References

1. Barnett, T.C., Patel, A.R. & Scott, J.R. A novel sortase, SrtC2, from *Streptococcus pyogenes* anchors a surface protein containing a QVPTGV motif to the cell wall. *J Bacteriol* **186**, 5865-5875 (2004).
2. Marraffini, L.A. & Schneewind, O. Anchor structure of staphylococcal surface proteins. V. Anchor structure of the sortase B substrate IsdC. *J Biol Chem* **280**, 16263-16271 (2005).
3. Mazmanian, S.K., Ton-That, H., Su, K. & Schneewind, O. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc Natl Acad Sci U S A* **99**, 2293-2298 (2002).
4. Maresso, A.W., Chapa, T.J. & Schneewind, O. Surface protein IsdC and Sortase B are required for heme-iron scavenging of *Bacillus anthracis*. *J Bacteriol* **188**, 8145-8152 (2006).
5. Gaspar, A.H. et al. *Bacillus anthracis* sortase A (SrtA) anchors LPXTG motif-containing surface proteins to the cell wall envelope. *J Bacteriol* **187**, 4646-4655 (2005).
6. Budzik, J.M., Oh, S.Y. & Schneewind, O. Cell wall anchor structure of BcpA pili in *Bacillus anthracis*. *J Biol Chem* **283**, 36676-36686 (2008).
7. Weiner, E.M., Robson, S., Marohn, M. & Clubb, R.T. The Sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J Biol Chem* **285**, 23433-23443 (2010).
8. Janulczyk, R. & Rasmussen, M. Improved pattern for genome-based screening identifies novel cell wall-attached proteins in gram-positive bacteria. *Infect Immun* **69**, 4019-4026 (2001).
9. Liew, C.K. et al. Localization and mutagenesis of the sorting signal binding site on sortase A from *Staphylococcus aureus*. *FEBS Lett* **571**, 221-226 (2004).
10. Bentley, M.L., Gaweska, H., Kielec, J.M. & McCafferty, D.G. Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J Biol Chem* **282**, 6571-6581 (2007).
11. Bentley, M.L., Lamb, E.C. & McCafferty, D.G. Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J Biol Chem* **283**, 14762-14771 (2008).
12. Anderson, E. et al. Transcytosis of NgCAM in epithelial cells reflects differential signal recognition on the endocytic and secretory pathways. *J Cell Biol* **170**, 595-605 (2005).
13. Feinberg, E.H. et al. GFP Reconstitution Across Synaptic Partners (GRASP) defines cell contacts and synapses in living nervous systems. *Neuron* **57**, 353-363 (2008).
14. Thyagarajan, A. & Ting, A.Y. Imaging activity-dependent regulation of neuroligin-neurexin interactions using trans-synaptic enzymatic biotinylation. *Cell* **143**, 456-469 (2010).

**Figure Legends**

**Figure 8.1. Photocrosslinking of membrane proteins**

Protein complexes can be defined by using sortase to site-specifically affix a probe bearing a photocrosslinker moiety onto a protein of interest in the rigorous absence of detergent (either in semi-intact cells or on the cell surface). Following crosslinking and immunoprecipitation, the covalent adduct can be identified by mass spectrometry.

**Figure 8.2. A system for reporting on transient cell-cell interactions**

To identify a cell-cell interactions, sortase is expressed as a fusion protein on the surface of one cell, while the interacting cell expresses a protein bearing an LPXTG motif. Addition of exogenous nucleophile will leave an indelible mark on the cell bearing the LPXTG motif and thus report on transient interactions.



**Figures**

**Figure 8.1.**

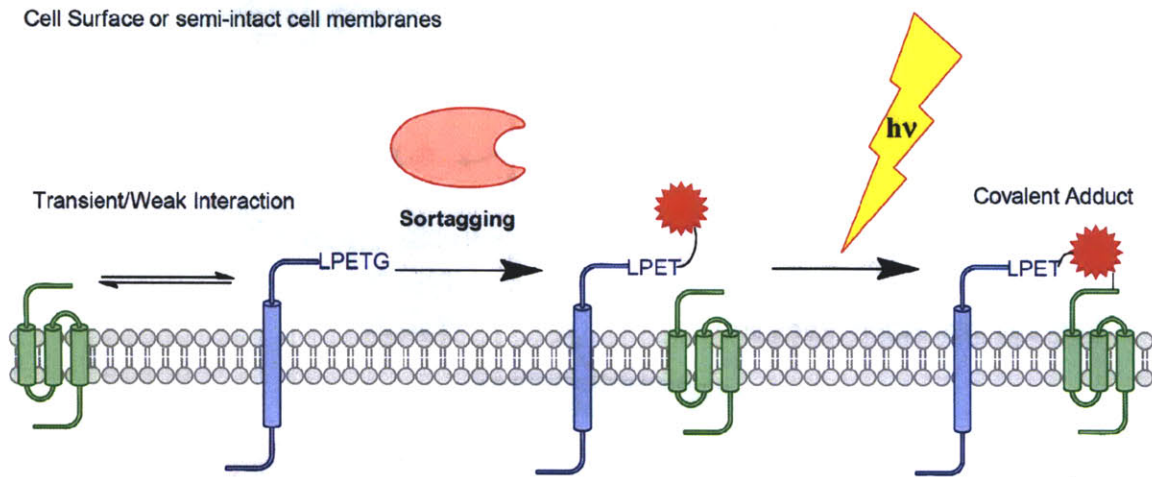
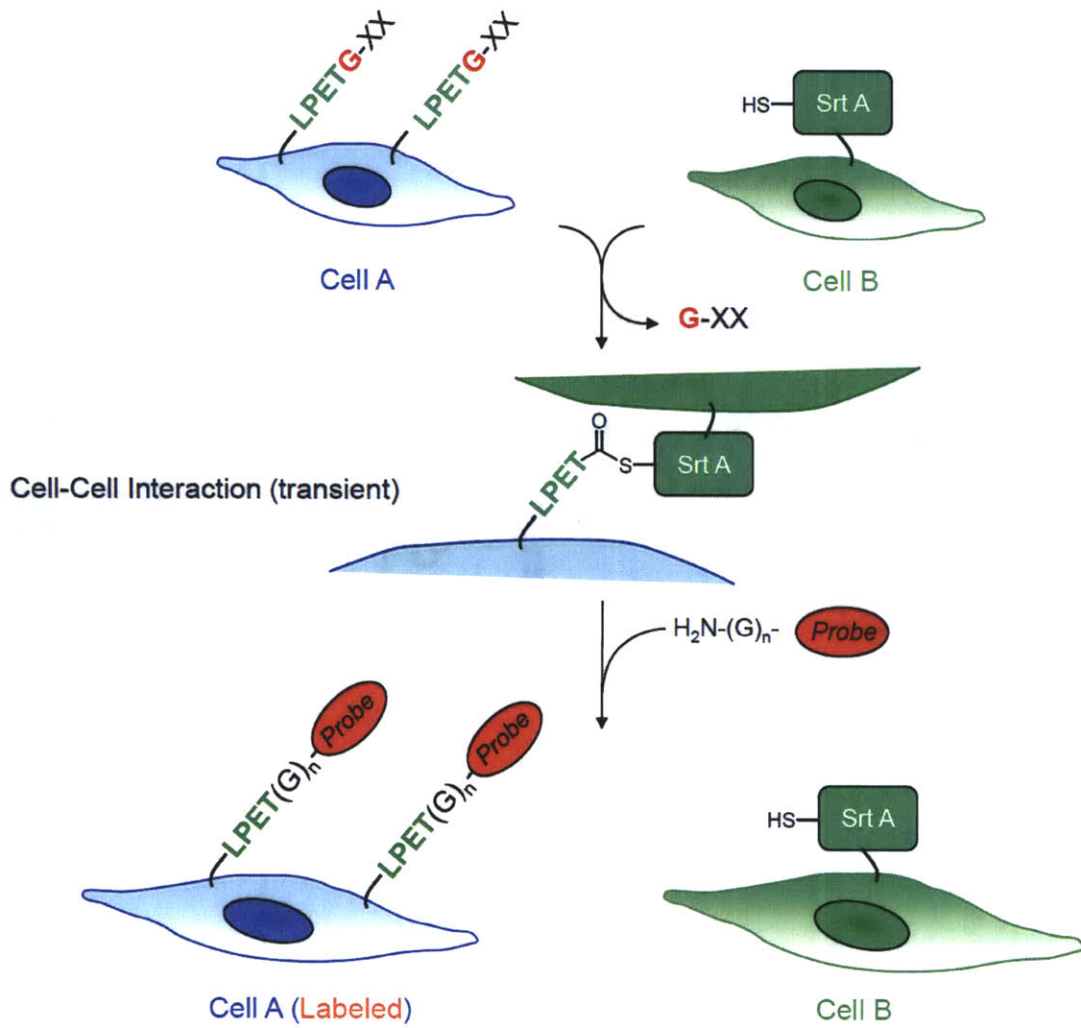


Figure 8.2



Print Article (English Version, reprinted with permission from *Angewandte Chemie International Edition*):

## Minireviews

H. L. Ploegh and M. W.-L. Popp

Protein Engineering

DOI: 10.1002/anie.201008267

# Making and Breaking Peptide Bonds: Protein Engineering Using Sortase

Maximilian Wei-Lin Popp and Hidde L. Ploegh\*

protein engineering · protein modifications · site specificity · sortase · transpeptidation

**S**ortases are a class of bacterial enzymes that possess transpeptidase activity. It is their ability to site-specifically break a peptide bond and then reform a new bond with an incoming nucleophile that makes sortase an attractive tool for protein engineering. This technique has been adopted for a range of applications, from chemistry-based to cell biology and technology. In this Minireview we provide a brief overview of the biology of sortase enzymes and current applications in protein engineering. We identify areas that lend themselves to further innovation and that suggest new applications.

## 1. Biological Function and Biochemistry of Sortase A (SrtA)

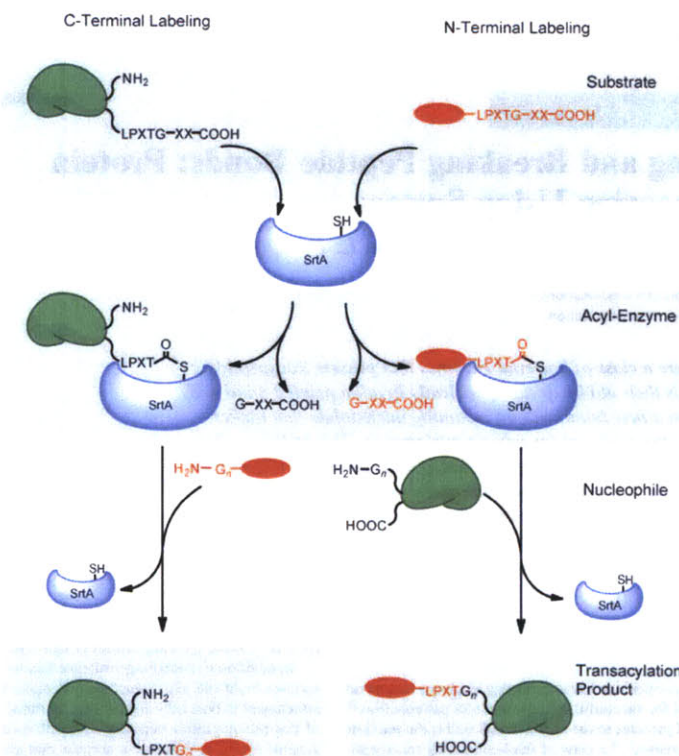
Many Gram-positive bacteria display virulence factors on their cell wall for successful colonization and pathogenesis.<sup>[1]</sup> Anchoring of proteins to the bacterial cell wall is the purview of sortase enzymes,<sup>[2–4]</sup> a class of thiol-containing transpeptidases. These enzymes recognize substrate proteins bearing a “sorting motif” (LPXTG in the case of *Staphylococcus aureus*) and harbor a catalytic cysteine residue that is used to cleave the peptide bond between the threonine and glycine residues within this pentapeptide.<sup>[5–7]</sup> Other sortases from different bacterial species use the same or similar recognition sequences.<sup>[8–10]</sup> A database of sortases and their substrates can be found on the internet.<sup>[11,12]</sup> This peptide-cleaving reaction initially yields a thioacyl intermediate,<sup>[13,14]</sup> in a fashion analogous to the mechanism used by cysteine proteases. Whereas cysteine proteases use water to resolve this intermediate and generate a hydrolysis product, sortase will accept the N terminus of an oligoglycine nucleophile, thereby resulting in the creation of a new peptide bond (Figure 1). In the course of the normal sorting reaction, the pentaglycine crossbridge in the lipid II cell wall precursor carries out the nucleophilic attack on the acyl-enzyme.<sup>[15]</sup> The cell-wall

precursor with its covalently attached protein is then incorporated into the growing peptidoglycan layer.

In addition to anchoring virulence factors to the cell wall, sortases build the pilus structure that many bacteria use for attachment to host cells and to form biofilms.<sup>[16–19]</sup> The details of this process differ between bacterial species,<sup>[20–24]</sup> but, in general terms, it involves a sortase that polymerizes pilin subunits bearing both a sorting signal and a nucleophilic ε-amine of a lysine residue in an internally located motif.<sup>[25]</sup> This protein–protein ligation reaction results in polymerization of the pilin subunits, but does not mediate anchoring of the growing polymer to the cell wall. This is the job of the housekeeping sortase, which accepts the lipid II precursor nucleophile.<sup>[26]</sup>

Sortases represent a bona fide drug target because of their central role in virulence.<sup>[27–31]</sup> Recombinant sortase lacking its transmembrane domain is readily produced in high yield.<sup>[6,32,33]</sup> For a detailed protocol for sortase production, see Ref. [34]. This has facilitated extensive structural<sup>[35–38]</sup> and biochemical<sup>[39–41]</sup> studies of the enzyme. The structure of sortase A from *S. aureus* consists of an eight-stranded β-barrel fold structure, termed the sortase fold, with a hydrophobic cleft formed by the β7 and β8 strands. This cleft is surrounded by the β3–β4, β2–β3, β6–β7, and β7–β8 loops (Figure 2). This cleft houses the catalytic cysteine residue (Cys184) and accommodates substrate binding. An additional structural feature of the *S. aureus* enzyme is a calcium binding site formed by the β3/β4 loop. The calcium ion binds to this site through coordination to a residue in the β6/β7 loop. This binding slows its motion, thereby allowing the substrate to bind and increasing the activity eightfold.<sup>[42]</sup> The biochemical details of the active site include a key histidine residue (H120) that can form a thiolate–imidazolium ion pair with the

[\*] M. W.-L. Popp, Prof. Dr. H. L. Ploegh  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge, MA 02142 (USA)  
and  
Department of Biology, Massachusetts Institute of Technology  
77 Massachusetts Avenue, Cambridge, MA 02142 (USA)  
E-mail: ploegh@wi.mit.edu



**Figure 1.** Site-specific C- and N-terminal labeling scheme using sortase A. C-Terminal labeling (left) and N-terminal labeling (right) proceed through a substrate-recognition step (top), followed by generation of a thioacyl intermediate (middle) and resolution of the acylated enzyme by an exogenously added nucleophile (bottom). See text for details.

catalytic cysteine residue.<sup>[43]</sup> It is the deprotonated form of the cysteine residue that is competent for catalysis. However, at a physiological pH value, the ionized forms of these key amino acids are in equilibrium with the neutral forms, and only a

small percentage (ca. 0.06%) of the total enzyme is catalytically competent at any given time.<sup>[44,45]</sup> The cysteine residue attacks the amide bond between the threonine and glycine residues in the sorting motif. The protonated imidazolium ion

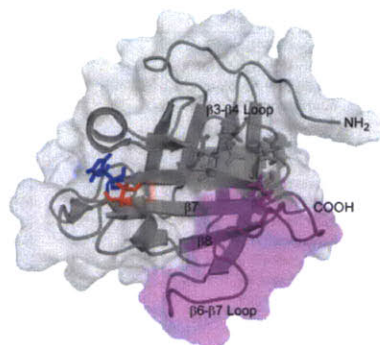


Hidde Ploegh received his BS in 1975 from Groningen University and his PhD from Leiden University, The Netherlands (1981), for his work in the laboratory of Jack Strominger at Harvard University (Cambridge, MA). From 1980 he was laboratory leader in Cologne, Germany. From there he moved to The Netherlands Cancer Institute (1984–1992), to MIT (1992–1997), Harvard Medical School (1997–2005), before ending up at Whitehead Institute/MIT, his present location. His current research focuses on the immune system and the creation of chemical tools with which to probe it.



Max Popp received his BS in Biology and BA in Chemistry from the University of Rochester in 2005. He is a PhD student at MIT in the research group of Hidde Ploegh, where his research has focused on establishing the sortase-mediated protein-labeling method.





**Figure 2.** Structure of sortase A from *Staphylococcus aureus* deduced by NMR spectroscopy (pdb code: 1IJA).<sup>[36]</sup> The active-site cysteine residue (Cys184) is in red and the active-site histidine residue (His120) is in blue. The  $\beta 7$  and  $\beta 8$  strands that form the floor of the active site are labeled and the  $\beta 6$ - $\beta 7$  loop involved in substrate recognition<sup>[31]</sup> is in purple. Residues that coordinate calcium are shown as sticks.<sup>[42]</sup>

acts as a general acid for the departing  $\alpha\text{NH}_2$  group from the glycine residue, and gives rise to an acylated form of sortase. An incoming glycine nucleophile is then deprotonated, attacks the thioester, and re-establishes an amide bond. If instead water attacks the acyl-enzyme intermediate, the reaction yields the dead-end hydrolysis product.<sup>[46]</sup>

## 2. Engineering of Bacterial Surfaces

The sortase-mediated system of anchoring proteins to the cell wall of Gram-positive bacteria was first exploited to decorate these microbes with heterologous proteins. Such experiments require the creation of a genetic fusion of the heterologous protein to the sorting motif. The heterologous protein is then expressed and directed to the surface through the normal cell-wall sorting pathway. In this manner, the enzyme alkaline phosphatase has been anchored to the cell wall of *Staphylococcus aureus*.<sup>[47]</sup> The E7 protein of Human papilloma virus 16 (HPV16) has been displayed on *Streptococcus gordonii*, a commensal microbe in the oral cavity,<sup>[48]</sup> and  $\alpha$ -amylase has been affixed to the peptidoglycan of *Bacillus subtilis*, helped by coexpression of the sortase gene from *Listeria monocytogenes*.<sup>[49]</sup> The peptidoglycan cell wall can even be decorated with non-natural entities (fluorescein, biotin, azide) by incubating dividing *S. aureus* cultures with chemical probes appended to the N terminus of an LPXTG peptide.<sup>[50]</sup> The incorporation of what are in essence N-terminal labeling probes (see Section 4) occurs through use of the endogenous sortase enzyme and anchors the exogenously provided probes onto available pentaglycine side chains of the cell wall.

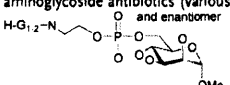
## 3. C-Terminal Labeling

The ability of sortase to recognize the sorting motif when transplanted onto recombinantly expressed proteins allows the site-specific incorporation of moieties and functional groups that cannot be encoded genetically (Figure 1). This method requires only that the LPXTG motif be solvent-exposed and usually results in high yields of the desired transpeptidation product. Indeed, many substrate proteins have now been labeled with probes bearing a wide range of functionalities, including biotin, fluorophores, cross-linkers, and multifunctional probes (Tables 1 and 2).<sup>[34]</sup> The labeling of recombinant proteins by sortase A requires no sophisticated synthetic chemistry; most of the probes are readily accessible by standard peptide synthesis, using off-the-shelf reagents. The production and folding of recombinant substrate proteins is not usually compromised by the presence of the small LPXTG tag. Since all transformations are carried out using sortase under physiological buffer conditions (pH, ionic strength, ionic requirements) on substrates whose proper folding and activity status can be ascertained prior to starting the reaction, loss of biological activity is rarely, if ever, observed for the final product. The ability to engage in a sortase-catalyzed transacylation appears to be determined solely by the accessibility and flexibility of the sorting motif. Intein-based protein engineering methods usually require that substrates first fold while fused to a protein-sized intein domain, which at times causes solubility issues.<sup>[51]</sup>

The utility of the sortase labeling method stems from the fact that the enzyme tolerates substrates unrelated in structure and sequence immediately upstream from the cleavage site. This property is not unexpected, given the role of sortase in anchoring a broad range of protein substrates to the cell wall. The substrate need not even be proteinaceous—peptide nucleic acids (PNAs) linked to the sortase cleavage site can be ligated to a glycine-linked cell-penetrating peptide (model amphipathic peptide, MAP) to yield active antisense PNA-peptide conjugates.<sup>[52]</sup> Likewise, the identity of the substituents C terminal to the glycine nucleophile do not seem to matter at all: D-amino acid containing peptides, folate, branched protein transduction domains,<sup>[53]</sup> and large polyethylene glycol chains<sup>[54]</sup> have all been attached using sortase. The cleavage site need not even be near the C terminus of the substrate protein. A sufficiently large solvent-exposed loop will suffice. This property has been exploited to investigate the contribution of a key loop in the deubiquitinating enzyme, ubiquitin C-terminal hydrolase 3 (UCHL3), to substrate binding and catalysis.<sup>[55]</sup> Since the cleavage site can be placed in a loop, it is possible to interrupt the connectivity of the protein backbone, while simultaneously installing a reporter moiety (biotin or fluorophore) to monitor the behavior of the cleaved enzyme in the presence of uncleaved, wild-type enzyme. This trait is likely to apply to many proteins whose conformation includes an exposed, flexible loop.

The C-terminal labeling technique is particularly useful for the study of type II membrane proteins embedded in the living mammalian cell membrane. Type II membrane proteins have C termini that are exposed to extracellular space and

**Table 1:** Examples of synthetic nucleophiles used in site-specific sortase A transpeptidation reactions.

Probe <sup>[a]</sup>	Labeling site	Property endowed	Reference(s)
1 H-G <sub>1</sub> K(biotin) L-OH	C terminal	biophysical handle	[32, 55]
2 H-G <sub>1</sub> K(ANP)K(biotin) L-OH	C terminal	biophysical handle/photocleavage	[32]
3 H-G <sub>1</sub> K(phenylazide)K(biotin) G-OH	C terminal	biophysical handle/photo-crosslinker	[32]
4 H-G <sub>1</sub> K(FITC)-NH <sub>2</sub>	C terminal	fluorescence	[32]
5 H-G <sub>1</sub> K(K(TAMRA))-NH <sub>2</sub>	C terminal	fluorescence	[32, 71]
6 H-G <sub>1</sub> YC(biotin)-NH <sub>2</sub>	C terminal	biophysical handle	[33]
7 H-G <sub>1</sub> YC(Alexa 488)-NH <sub>2</sub>	C terminal	fluorescence	[33]
8 H-AA-Ahx-K(K(TAMRA))-NH <sub>2</sub>	C terminal ( <i>S. pyogenes</i> )	fluorescence	[56]
9 H-G <sub>1</sub> K(C12-C24)-NH <sub>2</sub>	C terminal	lipidation	[64]
10 H-G <sub>1</sub> K(1-ad)-NH <sub>2</sub>	C terminal	hydrophobicity	[64]
11 H-G <sub>1</sub> WK(cholesterol)-NH <sub>2</sub>	C terminal	lipidation	[64]
12 D-Tat (1st residue is G)	C terminal	cell penetration	[53]
13 H-G <sub>1</sub> Y-PTD <sub>3</sub> -NH <sub>2</sub>	C terminal	cell penetration	[53]
14 (H <sub>2</sub> NRRQRRTSKLMKRAhx) <sub>2</sub> KYK(GG-NH <sub>2</sub> )-NH <sub>2</sub>	C terminal	cell penetration	[53]
15 H-G <sub>1</sub> K(folate)-NH <sub>2</sub>	C terminal	folic acid	[53]
16 H <sub>2</sub> N-PEG	C terminal	inert polymer	[54]
17 H-G <sub>1</sub> K(PEG)-OH	C terminal	inert polymer	[54]
18 H-G <sub>1</sub> -MAP-NH <sub>2</sub>	C terminal	cell penetration	[52]
19 aminoglycoside antibiotics (various and enantiomer)	C terminal	antibiotic	[60]
20 	C terminal	GPI mimic	[61]
21 GPI mimics based on 19 with trisaccharide cores	C terminal	GPI mimic	[62, 63]
22 biotin-PEG-VGLPETGG-NH <sub>2</sub>	N terminal	biophysical handle	[57]
23 Alexa 647-LPETGG-NH <sub>2</sub>	N terminal	fluorescence	[57]
24 Alexa 488-LPETGG-NH <sub>2</sub>	N terminal	fluorescence	[58]
25 biotin-LPRT-OMe	N terminal	biophysical handle	[56]
26 FITC-Ahx-LPRT-OMe	N terminal	fluorescence	[56]
27 FAM-LPETG-NH <sub>2</sub>	N terminal	fluorescence	[50]
28 biotin-GGLPETG-NH <sub>2</sub>	N terminal	biophysical handle	[50]
29 N <sub>3</sub> -ALPETG-NH <sub>2</sub>	N terminal	handle for bioorthogonal chemical reactions	[50]

[a] 1-Ad = 1-adamantyl, Ahx = aminohexanoic acid, FAM = carboxyfluorescein, FITC = fluorescein isothiocyanate, PEG = polyethylene glycol, TAMRA = carboxytetramethylrhodamine.

thus are excellent candidates for sortase-mediated labeling. Proteins with this type II topology have been particularly refractory to genetic fusion with fluorescent proteins. Placement of a fluorescent protein at the N terminus usually impedes cotranslational insertion of the type II membrane protein into the endoplasmic reticulum (ER), while C-terminal tagging with green fluorescent protein (GFP) places this bulky substituent close to the site of interaction with ligands of the type II membrane protein in question. The CD40 ligand protein (CD40L), influenza neuraminidase,<sup>[52]</sup> and osteoclast differentiation factor (ODF)<sup>[53]</sup> have all been labeled in live cells in this way.

#### 4. N-Terminal Labeling

Protein labeling at the N terminus can be accomplished simply by moving the placement of the sortase recognition element from the protein to the short peptide probe and by inclusion of a suitable number of glycine residues at the N terminus of the target protein (Figure 1). Both methyl ester mimetics of the sortase motif<sup>[56]</sup> as well as the complete LPXTG sortase recognition motif can be used as scaffolds for such probes.<sup>[57]</sup> Conceptually, this labeling technique is

analogous to the C-terminal labeling, except the acyl-enzyme intermediate is generated between sortase and the peptide probe, and the protein to be labeled bears several glycine residues at the N terminus, the  $\alpha$ NH<sub>2</sub> group of which serves as the nucleophile. This strategy was used to install fluorescent probes at the N terminus of membrane proteins in living mammalian cells after a clever initial unmasking step by sortase itself to expose the nucleophilic glycine.<sup>[57]</sup> This system was later used to install reporter fluorophores on the N terminus of the G-protein-coupled platelet-activating factor receptor (PAFR). This allowed the direct observation of the trafficking of the cell-surface-exposed pool after labeling. PFAFR receptors with key mutations were then shown to traffic aberrantly.<sup>[58]</sup> For both the C-terminal and N-terminal labeling of cell-surface proteins, the sortase technique allows access only to the cell-surface pool of the protein of interest. This is an advantage when the behavior of only the surface-exposed fraction of a particular protein is of interest. If ligand binding is restricted to the cell surface, then this is also usually the relevant fraction. Genetic fusions to fluorescent proteins, by their very nature report on the protein of interest from the moment of its genesis inside the cell and onwards. Although this trait comes with its own advantages, it may complicate the distinction between proteins in the course of their biosynthe-

## Minireviews

H. L. Ploegh and M. W.-L. Popp

**Table 2:** Examples of proteins labeled by sortase A transpeptidation.

Substrate	Solution/cell surface	Labeling site	Label(s) <sup>[a]</sup>	Reference(s)
H-2K <sup>b</sup>	solution	C terminal	1,2,3,4,5	[32]
CD154	cell surface	C terminal	1,5	[32]
neuramidase	cell surface	C terminal	1	[32]
ODF	cell surface	C terminal	6,7	[33]
Cre	solution	C terminal	5	[71]
UCHL3	solution	C terminal (loop)	1	[55]
p97	solution	C terminal	5	[71]
eGFP	solution	C terminal	9,10,11	[64]
GFP	solution	C terminal	13,14,15	[53]
PNA	solution	C terminal	18	[52]
eGFP	solution	C terminal	16,17	[54]
Mrp	solution	C terminal	19	[60]
YALPETGK	solution	C terminal	19	[60]
(His) <sub>3</sub> YALPETGKS	solution	C terminal	20	[61]
CD52 peptides	solution	C terminal	21	[62]
CD24	solution	C terminal	21	[62]
MUC1	solution	C terminal	21	[63]
LPETG <sub>3</sub> -ECFP-TM	cell surface	N terminal	22,23	[57]
LPETG <sub>3</sub> -PAFR	cell surface	N terminal	24	[58]
C <sub>12</sub> /C <sub>22</sub> -CTXB	solution	N terminal	25,26	[56]
G <sub>3</sub> -eGFP	solution	N terminal	26	[56]
G-UCHL3	solution	N terminal	26	[56]
<i>S. aureus</i> surface peptidoglycan	cell surface	N terminal	27,28,29	[50]
eGFP	solution	N and C terminal	26 and 8	[56]
UCHL3	solution	N and C terminal	26 and 8	[56]

[a] The numbers denote the probe identities from Table 1.

sis and the behavior of the mature, biologically relevant pool of protein. The sortase-based strategies should thus be viewed as a useful adjunct to the GFP-based methods, but with the added benefit of increased chemical flexibility.

### 5. Labeling at N and C Termini

It is possible to combine N-terminal and C-terminal labeling strategies by using sortases with distinct substrate specificity. The *Streptococcus pyogenes* enzyme<sup>[59]</sup> (SrtA<sub>Strep</sub>) recognizes and cleaves the LPXTA motif and accepts alanine-based nucleophiles. It also cleaves the SrtA<sub>Staph</sub> (Sortase A from *Staphylococcus aureus*) LPXTG motif, albeit with reduced efficiency. In contrast, the SrtA<sub>Staph</sub> enzyme does not cleave the LPXTA motif, and thus the two enzymes are orthogonal with respect to the LPXTA sequence. This property was exploited to label both termini of GFP and UCHL3 with different fluorophores. A masking strategy was used in which the N-terminal glycine residues needed for SrtA<sub>Staph</sub> labeling were exposed after proteolytic cleavage by thrombin. This step avoids protein oligomerization, likely to occur during the C-terminal labeling step with SrtA<sub>Strep</sub>.<sup>[56]</sup>

### 6. Post-Translational Modification Mimics

Sortase methods allow the production of homogeneous recombinant protein preparations that are modified with nongenetically templated post-translational modifications. Glycoproteins, normally elaborated by a complex set of

enzymatic events in the secretory pathway, can thus be constructed. LPXTG-tagged proteins and peptides can be modified with 6-aminohexose-based sugar nucleophiles, including aminoglycoside antibiotics and their analogues.<sup>[60]</sup> Glycosylphosphatidylinositol (GPI) anchors, normally attached at the C terminus of proteins, can be phenocopied by ligation of LPXTG peptides to synthetic glycine nucleophiles, which in turn are linked to the phosphoethanolamine moiety on a GPI derivative.<sup>[61]</sup> Various peptides (CD52 fragment, Mucin 1 peptides) and small proteins (CD24) have been attached to GPI mimics with trisaccharide cores.<sup>[62,63]</sup> Lipidation of proteins is yet another important post-translational modification that has been poorly studied because of the lack of tools available to obtain homogeneous preparations of lipoproteins. Sortase has been used to fill this void.<sup>[64]</sup> A glycine-based scaffold was modified with a panel of linear alkyl chains (C<sub>12</sub>–C<sub>24</sub>) as well as with cholesterol or adamantane, and then used to modify a suitably LPETG-tagged version of eGFP. These eGFP lipoproteins associated with the plasma membranes of living cells in a chain-length-dependent fashion (the optimum being a C<sub>22</sub> chain), from where they gained access to the endosomal compartment.

### 7. Piecemeal Assembly of Proteins, Protein Domains, and Peptides

Folded proteins with an exposed glycine residue at the N terminus may serve as nucleophiles for sortase labeling. Substrate proteins bearing the LPXTG motif can be fused to the incoming nucleophile protein, thereby creating large

transpeptidation products. By using independently folded proteins as substrates, it is possible to avoid many of the solubility issues that plague the expression of large genetically encoded fusions. This property was exploited to facilitate structural analysis by NMR spectroscopy, which typically requires highly concentrated protein preparations, thus making poor solubility a major obstacle. Sortase was used to attach an unlabeled, and hence NMR-invisible, solubility-enhancing tag based on the B1 immunoglobulin binding domain (G<sub>1</sub>-GB1) onto the C terminus of the Vav SRC homology 3 domain (SH3), a domain that is nearly insoluble by itself at pH 7.<sup>[65]</sup> The structure of the attached <sup>13</sup>C/<sup>15</sup>N-labeled Vav SH3 domain was then resolved by NMR spectroscopy, without confounding signals from the solubility-enhancing tag. Segmental labeling of the MecA protein of *Bacillus subtilis* using sortase for NMR spectroscopic study has also been reported.<sup>[66]</sup> A versatile panel of immunodetection reagents has been created using sortase. Protein–protein ligations were carried out between an Fc binding module (ZZ domain) and several detection enzymes (alkaline phosphatase, luciferase, glucose oxidase) using sortase.<sup>[67]</sup> Mucin glycopeptides that contain both N- and O-linked glycans were synthesized with the help of sortase. Separate peptides bearing either O- or N-linked glycans were constructed by a combination of chemical synthesis and elaboration of the glycan structure by enzymatic synthesis. These glycopeptides were then stitched together using sortase to yield a stereochemically homogeneous preparation.<sup>[68,69]</sup> Sortase has been applied to the construction of G-protein-coupled receptor (GPCR) mimics through a combined recombinant, enzymatic, and chemical synthesis (CRECS) strategy.<sup>[70]</sup> GPCRs are proteins with seven transmembrane regions (7-pass transmembrane proteins) that use three extracellular loops as well as the extracellular N-terminal segment to bind their ligands. To mimic this arrangement, three loops were made synthetically, cyclized by native chemical ligation, and appended to a triglycine-linked peptide scaffold. Then, the N terminus of GPCR, fused to the sortase cleavage site, was recombinantly expressed in *E. coli* and attached to the scaffold through sortase-mediated ligation. These elegantly engineered soluble mimics should allow the systematic characterization of the contributions of each region to ligand binding, and represent a true marriage between what can be accomplished through chemical synthesis and molecular biology.

A unique variant of the protein–protein ligation occurs when the LPXTG motif and N-terminal glycine residues are both present in the same construct. If both units are sufficiently close in space in the folded protein, the N terminus can form a peptide bond with the sortase recognition element, thereby resulting in a stable, circular transpeptidation product.<sup>[54,71]</sup> Circular proteins have useful biochemical properties. They are resistant to aggregation, require more energy for denaturation, and, since they lack exposed termini, are resistant in their native form to exoprotease attack.<sup>[72–76]</sup>

## 8. Anchoring to Solid Surfaces

Covalent immobilization of proteins onto solid supports has been accomplished by sortase. A major advantage of the method is that the specificity of the enzyme enables proteins to be immobilized uniformly and in a defined orientation on the solid surface for subsequent exposure to the analyte of interest. Stringent wash conditions can be employed because of the stable amide bond that links the protein to the surface, as was done by covalently attaching GFP to glycine-derivatized polystyrene beads.<sup>[41]</sup> The attachment of adhesion proteins from Gram-positive bacteria to fluorescent glycine-derivatized polystyrene beads was done in a similar manner.<sup>[77]</sup> The anchoring of GFP and Tus proteins to glycidyl methacrylate beads derivatized with oligoglycine, as well as to glycine-modified agarose resin (Affi-Gel) and glycine modified aminosilane coated glass slides, has been achieved.<sup>[78]</sup> The directional anchoring of proteins onto triglycine-modified carboxymethylated dextran-based Biacore sensor chips for use in surface plasmon resonance has also been accomplished.<sup>[79]</sup> Recombinant fibronectin-binding protein (rFba-LPETG) from group A streptococcus (GAS) was anchored in this manner, which then allowed the measurement of binding of human factor H to the immobilized protein. With an aim to develop the reagents needed for chemoenzymatic synthesis of glycoconjugates, immobilized  $\beta$ -1,4-galactosyltransferase (rhGalT) and *Helicobacter pylori*  $\alpha$ -1,3-fucosyltransferase (rHFucT) were covalently attached to alkylamine-sepharose beads. These enzymes are both active and reusable when directionally anchored to the solid phase.<sup>[80]</sup>

## 9. Protein Expression and Purification

Genetic fusions between sortase and a protein of interest have been constructed for the purposes of protein purification. A linear fusion between hexahistidine-tagged sortase, the LPETG tag, and the protein of interest is first purified by nickel nitrilotriacetic acid immobilized metal affinity chromatography (Ni-NTA IMAC) and then cleaved off of the resin by addition of Ca<sup>2+</sup> and triglycine to yield highly pure protein with one additional glycine residue<sup>[81]</sup> (Figure 3). This method was adapted for protein production in a wheat germ cell-free translation system, with the goal of creating a general purification method that can be used in automated, high-throughput protein production. In this version, a biotin acceptor peptide (a 15 amino acid peptide that is the target of *E. coli* biotin ligase) replaces the hexahistidine tag, and the proteins are purified with streptavidin resin in the presence of calcium chelators to prevent premature cleavage.<sup>[82]</sup> Both methods yield the target protein with one extra glycine residue at the N terminus, a configuration that is poised for N-terminal labeling by sortase if desired (Figure 3).

Sortase A from *S. aureus* is an extremely soluble enzyme that can be produced in high yield (>40 mg L<sup>-1</sup> of culture). This property has been exploited to enhance the solubility of proteins of interest by fusion to a version of sortase lacking the catalytic cysteine.<sup>[83]</sup>





**Figure 3.** Protein purification using sortase A. Recombinant expressed proteins are produced as fusion proteins containing either a hexahistidine tag (top)<sup>[81]</sup> or a biotin acceptor peptide (bottom)<sup>[82]</sup> followed by the catalytic core of sortase, the LPXTG tag, and the protein of interest. Addition of  $\text{Ca}^{2+}$  ions and oligoglycine to the immobilized fusion protein stimulates sortase activity. The protein of interest is released as a purified preparation with one additional N-terminal glycine.

### 10. Outlook

The diverse microbial world includes multiple sortase-type enzymes, with several distinct cleavage site preferences and nucleophile specificities.<sup>[84]</sup> The pilin-building sortases, for example, accept the  $\epsilon$ -amino group of a lysine residue within the YPKN sequence as a nucleophile, whereas sortase A from *S. aureus* will accept the N terminus of glycine extensions. Sortase B (SrtB) from *S. aureus* cleaves the NPQTN sequence,<sup>[85,86]</sup> while SrtB from *Bacillus anthracis* cleaves the NPKTG sequence<sup>[87]</sup> and accepts *meso*-diaminopimelic acid as a nucleophile.<sup>[88–90]</sup> Although the use of distinct sortases for protein labeling is in its infancy, there is clearly great untapped potential in these other sortases if the recognition requirements and reaction conditions can be more clearly specified. With the advent of many directed protein evolution techniques, it may be possible to improve on what nature has provided. In its usual biological context, the housekeeping sortase anchors to the cell wall a set of substrates that are structurally distinct;<sup>[91]</sup> the one feature required for attack by sortase seems to be the LPXTG pentapeptide. This, coupled with the fact that substrate recognition is primarily a function of the  $\beta 6$ - $\beta 7$  loop,<sup>[92–94]</sup> implies that it should be possible to evolve versions of sortase that will attack fully orthogonal peptide sequence motifs. Given the diversity of sortases that recognize different anchor structures,<sup>[88,89]</sup> it should be possible to evolve enzymes that will accept nucleophiles other than those that occur naturally. By combining sortases of distinct specificity, one can envision orthogonal labeling of different proteins in a one-pot reaction—or its biological equivalent—inserted into the same cellular environment. The kinetics of the sortase reaction are lackluster at best, most likely owing to the active-site equilibrium which favors the inactive, non-ionized cysteine residue. Although this problem is easily overcome by adding large quantities of sortase to the reaction (and thus adding a larger total amount of catalytically competent enzyme), it would still be beneficial to improve the overall kinetics of transpeptidation by *S. aureus* SrtA by using directed evolution.

The pilus-building sortases which polymerize subunits by creating isopeptide bonds remain to be exploited. These enzymes may be of particular interest, since they evolved to

establish covalent protein–protein linkages post-translationally, and perhaps this trait could be harnessed to perform similar protein–protein ligations in vitro, by exploiting exposed lysine residues internal to the protein of interest as the incoming nucleophile. Given that isopeptide bond formation is featured in many biological processes, for example in the ubiquitin system, the ability to make homogeneous, site-specific isopeptide linkages would be tremendously useful.

An intriguing question is how to overcome the limitations that prevent intracellular labeling by sortase. It is possible to express sortase inside of both living bacterial and eukaryotic cells—the enzyme folds properly and is active.<sup>[95]</sup> The *S. pyogenes* enzyme, which lacks the requirement for divalent cations, is active in compartments devoid of calcium, such as the cytosol. Similarly, substrates that bear an exposed LPXTG motif are easily expressed, and depending on the protein, retain their normal localization and/or function. When the nucleophile for transpeptidation is genetically encoded as part of the substrate protein, as in the case of circularization or protein–protein ligation reactions, transpeptidation proceeds readily,<sup>[95]</sup> and yields circular proteins intracellularly in yeast, mammalian cells, and prokaryotes. The limiting step for simple C-terminal labeling is thus delivery of an exogenous nucleophile to the cell interior in quantities that permit biochemical analysis of the transpeptidation product. These limitations might be overcome by attaching one of the many protein transduction sequences derived from the human immunodeficiency virus transactivator of transcription protein (HIV TAT) or other sources to the labeled nucleophile of interest, although this possibility remains to be achieved in practice and fails to address the fate of any unincorporated cytoplasmically delivered nucleophile. Alternatively, it might be possible to transiently mask the  $\alpha$ -amine group and other groups that preclude delivery to the cytosol through installation of chemical entities that can be removed by esterases, photochemically, or in the reducing environment of the cytoplasm, thus making probe delivery irreversible.

A key advantage of the sortase labeling method is the relative ease with which the nucleophilic probes are synthesized. Most probes are simple C-terminal appendages of glycine, made on a solid phase, with the reporter either incorporated while the peptide is on the resin or in solution by standard chemical coupling techniques. The literature shows that a diverse array of such probes will all likely react as nucleophiles with a given target in the same manner. This makes it possible to create libraries of bioconjugates without the need to reoptimize ligation conditions for each conjugate individually. Cell- and tissue-specific proteinaceous delivery vehicles for small interfering RNA (siRNA) libraries and antibody–drug library conjugates are all within the realm of possibility. An added benefit of the sortase method is the tight control over the stoichiometry of labeling: this should greatly facilitate quantitative studies of the effects of such bioconjugate libraries.

In addition to the strict control over the stoichiometry of ligation products, sortase allows control over the orientation between ligation partners, and this could be useful in creating protein–protein conjugates. The orientation of the subunits can be reversed by switching the location of the N-terminal

glycine residues on one unit for the sortase motif on the other partner protein. By using cleverly designed probes bearing handles for bioorthogonal reactions (for example, reactions between an azide and an alkyne, an azide and a strained cyclooctyne, or an aldehyde and an aminoxy compound), it should be possible to construct protein pairs that are covalently linked in non-natural topologies (N–N or C–C). Alternatively, two-step transacylation procedures using the semiorthogonal SrtA<sub>Suph</sub> and SrtA<sub>Strep</sub> enzymes along with the appropriate peptide probes in N–N or C–C linkages can be employed. The sortase method thus provides access to protein linkages not found in nature, and one area where this can be applied is in making engineered ligands for studying cell signaling. It would be possible to create chimeric, folded protein ligands that constrain binding and heterodimerization of their cognate receptors. It is clear that the ease of use and flexibility of the sortase labeling method will provide exciting new avenues of research.

Received: December 30, 2010

Published online: ■■■ ■■■ ■■■■

- [1] L. A. Marraffini, A. C. Dedent, O. Schneewind, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 192.
- [2] S. K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, *Science* **1999**, *285*, 760.
- [3] S. K. Mazmanian, H. Ton-That, O. Schneewind, *Mol. Microbiol.* **2001**, *40*, 1049.
- [4] S. D. Zink, D. L. Burns, *Infect. Immun.* **2005**, *73*, 5222.
- [5] H. Ton-That, O. Schneewind, *J. Biol. Chem.* **1999**, *274*, 24316.
- [6] H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12424.
- [7] R. G. Kruger, B. Otvos, B. A. Frankel, M. Bentley, P. Dostal, D. G. McCafferty, *Biochemistry* **2004**, *43*, 1541.
- [8] S. Dramsi, P. Trieu-Cuot, H. Bierné, *Res. Microbiol.* **2005**, *156*, 289.
- [9] M. G. Pucciarelli, E. Calvo, C. Sabet, H. Bierné, P. Cossart, F. Garcia-del Portillo, *Proteomics* **2005**, *5*, 4808.
- [10] D. Comfort, R. T. Clubb, *Infect. Immun.* **2004**, *72*, 2710.
- [11] a) Internet address: [http://bamics3.cmbi.kun.nl/cgi-bin/jos/sortase\\_substrates/index.py](http://bamics3.cmbi.kun.nl/cgi-bin/jos/sortase_substrates/index.py); b) J. Boekhorst, M. W. de Been, M. Kleerebezem, R. J. Siezen, *J. Bacteriol.* **2005**, *187*, 4928.
- [12] M. J. Pallen, A. C. Lam, M. Antonio, K. Dunbar, *Trends Microbiol.* **2001**, *9*, 97.
- [13] H. Ton-That, S. K. Mazmanian, K. F. Faull, O. Schneewind, *J. Biol. Chem.* **2000**, *275*, 9876.
- [14] A. Aulabaugh, W. Ding, B. Kapoor, K. Tabei, L. Alksne, R. Dushin, T. Zatz, G. Ellestad, X. Huang, *Anal. Biochem.* **2007**, *360*, 14.
- [15] A. M. Perry, H. Ton-That, S. K. Mazmanian, O. Schneewind, *J. Biol. Chem.* **2002**, *277*, 16241.
- [16] H. Ton-That, O. Schneewind, *Trends Microbiol.* **2004**, *12*, 228.
- [17] J. R. Scott, D. Zahner, *Mol. Microbiol.* **2006**, *62*, 320.
- [18] A. Mandlik, A. Swierczynski, A. Das, H. Ton-That, *Trends Microbiol.* **2008**, *16*, 33.
- [19] T. Proft, E. N. Baker, *Cell. Mol. Life Sci.* **2009**, *66*, 613.
- [20] H. Ton-That, O. Schneewind, *Mol. Microbiol.* **2003**, *50*, 1429.
- [21] H. Ton-That, L. A. Marraffini, O. Schneewind, *Mol. Microbiol.* **2004**, *53*, 251.
- [22] C. Manzano, C. Contreras-Martel, L. El Mortaji, T. Izore, D. Fenel, T. Vermet, G. Schoehn, A. M. Di Guilmi, A. Dessen, *Structure* **2008**, *16*, 1838.
- [23] F. Neiers, C. Madhurantakam, S. Falker, C. Manzano, A. Dessen, S. Normark, B. Henriques-Normark, A. Achour, *J. Mol. Biol.* **2009**, *393*, 704.
- [24] C. Manzano, T. Izore, V. Job, A. M. Di Guilmi, A. Dessen, *Biochemistry* **2009**, *48*, 10549.
- [25] J. M. Budzik, L. A. Marraffini, P. Souda, J. P. Whitelegge, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10215.
- [26] A. Mandlik, A. Das, H. Ton-That, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14147.
- [27] S. K. Mazmanian, G. Liu, E. R. Jensen, E. Lenoy, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5510.
- [28] G. K. Paterson, T. J. Mitchell, *Trends Microbiol.* **2004**, *12*, 89.
- [29] A. W. Maresco, O. Schneewind, *Pharmacol. Rev.* **2008**, *60*, 128.
- [30] N. Surce, M. E. Jung, R. T. Clubb, *Mini-Rev. Med. Chem.* **2007**, *7*, 991.
- [31] G. K. Paterson, T. J. Mitchell, *Microbes Infect.* **2006**, *8*, 145.
- [32] M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, *Nat. Chem. Biol.* **2007**, *3*, 707.
- [33] T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, *ChemBioChem* **2008**, *9*, 802.
- [34] M. W. Popp, J. M. Antos, H. L. Ploegh, *Curr. Protoc. Protein Sci.* **2009**, Chap. 15, Unit 153.
- [35] U. Ilangovan, J. Iwahara, H. Ton-That, O. Schneewind, R. T. Clubb, *J. Biomol. NMR* **2001**, *19*, 379.
- [36] U. Ilangovan, H. Ton-That, J. Iwahara, O. Schneewind, R. T. Clubb, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6056.
- [37] Y. Zong, T. W. Bice, H. Ton-That, O. Schneewind, S. V. Narayana, *J. Biol. Chem.* **2004**, *279*, 31383.
- [38] N. Surce, C. K. Liew, V. A. Villareal, W. Thieu, E. A. Fadeev, J. J. Clemens, M. E. Jung, R. T. Clubb, *J. Biol. Chem.* **2009**, *284*, 24465.
- [39] R. G. Kruger, P. Dostal, D. G. McCafferty, *Anal. Biochem.* **2004**, *326*, 42.
- [40] L. A. Marraffini, H. Ton-That, Y. Zong, S. V. Narayana, O. Schneewind, *J. Biol. Chem.* **2004**, *279*, 37763.
- [41] B. A. Frankel, Y. Tong, M. L. Bentley, M. C. Fitzgerald, D. G. McCafferty, *Biochemistry* **2007**, *46*, 7269.
- [42] M. T. Naik, N. Surce, U. Ilangovan, C. K. Liew, W. Thieu, D. O. Campbell, J. J. Clemens, M. E. Jung, R. T. Clubb, *J. Biol. Chem.* **2006**, *281*, 1817.
- [43] H. Ton-That, S. K. Mazmanian, L. Alksne, O. Schneewind, *J. Biol. Chem.* **2002**, *277*, 7447.
- [44] K. M. Connolly, B. T. Smith, R. Pilpa, U. Ilangovan, M. E. Jung, R. T. Clubb, *J. Biol. Chem.* **2003**, *278*, 34061.
- [45] B. A. Frankel, R. G. Kruger, D. E. Robinson, N. L. Kelleher, D. G. McCafferty, *Biochemistry* **2005**, *44*, 11188.
- [46] X. Huang, A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei, G. Ellestad, *Biochemistry* **2003**, *42*, 11307.
- [47] O. Schneewind, P. Model, V. A. Fischetti, *Cell* **1992**, *70*, 267.
- [48] G. Pozzi, M. Contorni, M. R. Oggioni, R. Manganelli, M. Tommasino, F. Cavalieri, V. A. Fischetti, *Infect. Immun.* **1992**, *60*, 1902.
- [49] H. D. Nguyen, W. Schumann, *J. Biotechnol.* **2006**, *122*, 473.
- [50] J. W. Nelson, A. G. Chamessian, P. J. McEnaney, R. P. Murelli, B. I. Kazmierczak, D. A. Spiegel, *ACS Chem. Biol.* **2010**, *5*, 1147.
- [51] M. Vila-Perelló, T. W. Muir, *Cell* **2010**, *143*, 191.
- [52] S. Pritz, Y. Wolf, O. Kraetke, J. Klose, M. Bienert, M. Beyermann, *J. Org. Chem.* **2007**, *72*, 3909.
- [53] H. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670.
- [54] R. Parthasarathy, S. Subramanian, E. T. Boder, *Bioconjugate Chem.* **2007**, *18*, 469.
- [55] M. W. Popp, K. Artavanis-Tsakonas, H. L. Ploegh, *J. Biol. Chem.* **2009**, *284*, 3593.

- [56] J. M. Antos, G. L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W. Popp, H. L. Ploegh, *J. Am. Chem. Soc.* **2009**, *131*, 10800.
- [57] T. Yamamoto, T. Nagamune, *Chem. Commun.* **2009**, 1022.
- [58] N. Hirota, D. Yasuda, T. Hashidate, T. Yamamoto, S. Yamaguchi, T. Nagamune, T. Nagase, T. Shimizu, M. Nakamura, *J. Biol. Chem.* **2010**, *285*, 5931.
- [59] P. R. Race, M. L. Bentley, J. A. Melvin, A. Crow, R. K. Hughes, W. D. Smith, R. B. Sessions, M. A. Kehoe, D. G. McCafferty, M. J. Banfield, *J. Biol. Chem.* **2009**, *284*, 6924.
- [60] S. Samantary, U. Marathe, S. Dasgupta, V. K. Nandicoori, R. P. Roy, *J. Am. Chem. Soc.* **2008**, *130*, 2132.
- [61] X. Guo, Q. Wang, B. M. Swarts, Z. Guo, *J. Am. Chem. Soc.* **2009**, *131*, 9878.
- [62] Z. Wu, X. Guo, Q. Wang, B. M. Swarts, Z. Guo, *J. Am. Chem. Soc.* **2010**, *132*, 1567.
- [63] Z. Wu, X. Guo, Z. Guo, *Chem. Commun.* **2010**, 46, 5773.
- [64] J. M. Antos, G. M. Miller, G. M. Grotenbreg, H. L. Ploegh, *J. Am. Chem. Soc.* **2008**, *130*, 16338.
- [65] Y. Kobashigawa, H. Kumeta, K. Ogura, F. Inagaki, *J. Biomol. NMR* **2009**, *43*, 145.
- [66] M. A. Refaei, A. Combs, D. J. Kojetin, J. Cavanagh, C. Caperelli, M. Rance, J. Sapitro, P. Tsang, *J. Biomol. NMR* **2011**, *49*, 3.
- [67] T. Sakamoto, S. Sawamoto, T. Tanaka, H. Fukuda, A. Kondo, *Bioconjugate Chem.* **2010**, *21*, 2227.
- [68] T. Matsushita, S. Nishimura, *Methods Enzymol.* **2010**, *478*, 485.
- [69] T. Matsushita, R. Sadamoto, N. Ohyabu, H. Nakata, M. Fumoto, N. Fujitani, Y. Takegawa, T. Sakamoto, M. Kuroguchi, H. Hinou, H. Shimizu, T. Ito, K. Naruchi, H. Togame, H. Takemoto, H. Kondo, S. Nishimura, *Biochemistry* **2009**, *48*, 11117.
- [70] S. Pritz, O. Kraetke, A. Klose, J. Klose, S. Rothmund, K. Fechner, M. Bienert, M. Beyermann, *Angew. Chem.* **2006**, *120*, 3698; *Angew. Chem. Int. Ed.* **2006**, *47*, 3642.
- [71] J. M. Antos, M. W. Popp, R. Ernst, G. L. Chew, E. Spooner, H. L. Ploegh, *J. Biol. Chem.* **2009**, *284*, 16028.
- [72] R. J. Clark, H. Fischer, L. Dempster, N. L. Daly, K. J. Rosengren, S. T. Nevin, F. A. Meunier, D. J. Adams, D. J. Craik, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13767.
- [73] R. J. Clark, J. Jensen, S. T. Nevin, B. P. Callaghan, D. J. Adams, D. J. Craik, *Angew. Chem.* **2010**, *122*, 6695; *Angew. Chem. Int. Ed.* **2010**, *49*, 6545.
- [74] M. Trabi, D. J. Craik, *Trends Biochem. Sci.* **2002**, *27*, 132.
- [75] D. J. Craik, *Science* **2006**, *311*, 1563.
- [76] A. S. Andersen, E. Palmqvist, S. Bang, A. C. Shaw, F. Hubalek, U. Ribel, T. Hoeg-Jensen, *J. Pept. Sci.* **2010**, *16*, 473.
- [77] S. Wu, T. Proft, *Biotechnol. Lett.* **2010**, *32*, 1713.
- [78] L. Chan, H. F. Cross, J. K. She, G. Cavalli, H. F. Martins, C. Neylon, *PLoS One* **2007**, *2*, e1164.
- [79] F. Clow, J. D. Fraser, T. Proft, *Biotechnol. Lett.* **2008**, *30*, 1603.
- [80] T. Ito, R. Sadamoto, K. Naruchi, H. Togame, H. Takemoto, H. Kondo, S. Nishimura, *Biochemistry* **2010**, *49*, 2604.
- [81] H. Mao, *Protein Expression Purif.* **2004**, *37*, 253.
- [82] S. Matsunaga, K. Matsuoka, K. Shimizu, Y. Endo, T. Sawasaki, *BMC Biotechnol.* **2010**, *10*, 42.
- [83] J. Caswell, P. Snoddy, D. McMeel, R. J. Buick, C. J. Scott, *Protein Expression Purif.* **2010**, *70*, 143.
- [84] T. C. Barnett, A. R. Patel, J. R. Scott, *J. Bacteriol.* **2004**, *186*, 5865.
- [85] L. A. Marraffini, O. Schneewind, *J. Biol. Chem.* **2005**, *280*, 16263.
- [86] S. K. Mazmanian, H. Ton-That, K. Su, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2293.
- [87] A. W. Maresco, T. J. Chapa, O. Schneewind, *J. Bacteriol.* **2006**, *188*, 8145.
- [88] A. H. Gaspar, L. A. Marraffini, E. M. Glass, K. L. Debord, H. Ton-That, O. Schneewind, *J. Bacteriol.* **2005**, *187*, 4646.
- [89] J. M. Budzik, S. Y. Oh, O. Schneewind, *J. Biol. Chem.* **2008**, *283*, 36676.
- [90] E. M. Weiner, S. Robson, M. Marohn, R. T. Clubb, *J. Biol. Chem.* **2010**, *285*, 23433.
- [91] R. Janulczyk, M. Rasmussen, *Infect. Immun.* **2001**, *69*, 4019.
- [92] C. K. Liew, B. T. Smith, R. Pilpa, N. Suree, U. Ilangovan, K. M. Connolly, M. E. Jung, R. T. Clubb, *FEBS Lett.* **2004**, *571*, 221.
- [93] M. L. Bentley, H. Gaweska, J. M. Kielec, D. G. McCafferty, *J. Biol. Chem.* **2007**, *282*, 6571.
- [94] M. L. Bentley, E. C. Lamb, D. G. McCafferty, *J. Biol. Chem.* **2008**, *283*, 14762.
- [95] K. Strijbis, M. W. Popp, unpublished results.



Print Article (German Version, reprinted with permission from Angewandte Chemie):

## Kurzaufsätze

H. L. Ploegh und M. W.-L. Popp

Protein-Engineering

DOI: 10.1002/ange.201008267

Bilden und Brechen von Peptidbindungen:  
Protein-Engineering mithilfe von Sortase

Maximilian Wei-Lin Popp und Hidde L. Ploegh\*

Protein-Engineering · Proteinmodifikationen ·  
Sortasen · Ortsspezifität · Transpeptidierung

**S**ortasen gehören zu einer Klasse bakterieller Enzyme, die Transpeptidaseaktivität aufweisen. Die Fähigkeit, Peptidbindungen ortsspezifisch zu brechen und neue Bindungen mit einem angreifenden Nukleophil zu bilden, macht Sortase zu einem hervorragenden Hilfsmittel für das Protein-Engineering. Diese Methode wurde für eine Reihe von Anwendungen angepasst, die von der Chemie über die Zellbiologie bis hin zu technischen Gebieten reichen. Wir wollen hier einen kurzen Überblick über die Biologie der Sortaseenzyme und ihre Anwendungen im Protein-Engineering geben, auf Gebiete für zukünftige Innovationen hinweisen und neue Anwendungen vorschlagen.

1. Biologische Funktion und Biochemie von  
Sortase A (SrtA)

Viele Gram-positive Bakterien haben Virulenzfaktoren in ihrer Zellwand, die eine erfolgreiche Kolonisierung und Pathogenese ermöglichen.<sup>[1]</sup> Sortaseenzyme sind eine Klasse von Transpeptidasen, die Thiolgruppen enthalten, und haben die Fähigkeit, Proteine in der bakteriellen Zellwand zu verankern.<sup>[2–4]</sup> Diese Enzyme erkennen Substratproteine, die ein „Sortierungsmotiv“ (LPXTG im Fall von *Staphylococcus aureus*) enthalten, und tragen einen katalytischen Cysteinrest, um die Peptidbindung zwischen Threonin und Glycin in diesem Pentapeptid zu spalten.<sup>[5–7]</sup> Andere Sortasen verschiedener Bakterienarten nutzen die gleiche oder eine ähnliche Erkennungssequenz.<sup>[8–10]</sup> Datensammlungen für Sortase und ihre Substrate sind im Internet zu finden.<sup>[11,12]</sup> Die von der Sortase katalysierte Reaktion führt zuerst zu einer Thioacylzwischenstufe,<sup>[13,14]</sup> analog dem Mechanismus von Cysteinproteasen. Wo Cysteinproteasen jedoch Wasser zur Hydrolyse der Zwischenstufe nutzen, akzeptiert Sortase den N-Terminus eines Oligoglycinnukleophils zur Bildung einer neuen Peptidbindung (Abbildung 1). Bei der natürlichen

Sortierungsreaktion bindet die Pentaglycinbrücke in der Lipid-II-Zellwandvorstufe das Acylenzym als Nukleophil.<sup>[15]</sup> Die Zellwandvorstufe mit ihrem kovalent gebundenen Protein wird anschließend in die wachsende Peptidoglycanschicht eingebaut.

Zusätzlich zur Verankerung von Virulenzfaktoren in der Zellwand bauen Sortasen auch die Pilusstruktur auf, die von vielen Bakterien genutzt wird, um sich an der Wirtszelle zu verankern und einen Biofilm zu bilden.<sup>[16–19]</sup> Die Details dieses Prozesses unterscheiden sich zwar für verschiedene Bakterienarten,<sup>[20–24]</sup> aber im Allgemeinen verknüpfen Sortasen Pilinmonomere, die sowohl ein Sortierungssignal als auch eine nukleophile Lysin-ε-Aminogruppe in einem internen Erkennungsmotiv enthalten.<sup>[25]</sup> Diese Protein-Protein-Verknüpfung führt zu einer Polymerisation der Pilinmonomere, vermittelt jedoch nicht die Verankerung an der Zellwand. Die Verankerung wird vielmehr von der „Housekeeping“-Sortase ausgeführt, die das Lipid-II-Vorstufennukleophil aufnimmt.<sup>[26]</sup>

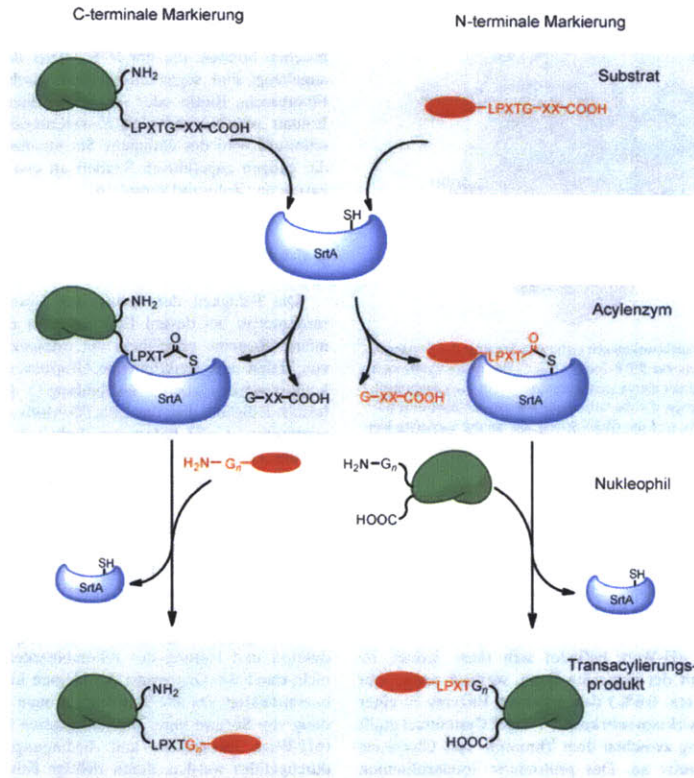
Wegen ihrer zentralen Bedeutung für die Virulenz gelten Sortasen als vielversprechendes Angriffsziel für Wirkstoffe.<sup>[27–31]</sup> Rekombinante Sortase, der die Transmembrandomäne fehlt, kann einfach und in hoher Ausbeute hergestellt werden.<sup>[6,32,33]</sup> Eine detaillierte Vorschrift für die Sortascherstellung, beschrieben in Lit. [34], hat ausführliche strukturelle<sup>[35–38]</sup> und biochemische Untersuchungen<sup>[39–41]</sup> des Enzyms ermöglicht. Die Struktur der Sortase A von *S. aureus* besteht aus einer achtsträngigen β-Barrel-Faltstruktur (genannt Sortasefaltstruktur) mit einer hydrophoben Spalte, die von den β7- und β8-Strängen gebildet wird. Diese Spalte ist von

\* M. W.-L. Popp, Prof. Dr. H. L. Ploegh  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge, MA 02142 (USA)  
und  
Department of Biology, Massachusetts Institute of Technology  
77 Massachusetts Avenue, Cambridge, MA 02142 (USA)  
E-Mail: ploegh@wi.mit.edu

## Chapter 8: Future Directions

Protein-Engineering mit Sortasen

Angewandte  
Chemie



**Abbildung 1.** Ortschafts-spezifische C- und N-terminale Markierung unter Verwendung von Sortase A. C-terminale (links) und N-terminale Markierung (rechts) erfolgen durch einen Substraterkennungsschritt (oben), gefolgt von der Bildung eines Thioacylintermediats (Mitte) und der Auflösung des acylierten Enzyms durch ein exogen zugeführtes Nukleophil (unten). Siehe Text für Details.

$\beta$ 3- $\beta$ 4-,  $\beta$ 2- $\beta$ 3-,  $\beta$ 6- $\beta$ 7- und  $\beta$ 7- $\beta$ 8-Strängen umgeben (Abbildung 2). Die Falte enthält den katalytischen Cysteinrest (Cys184) und die Substratbindestelle. Ein zusätzliches Strukturmerkmal des Enzyms von *S. aureus* ist eine Bindestelle für Calcium, die von der  $\beta$ 3/ $\beta$ 4-Schleife gebildet wird.

Das Calciumion, das an dieser Stelle bindet, wird von einem Rest in der  $\beta$ 6/ $\beta$ 7-Schleife koordiniert. Dadurch vermindert sich die Bewegungsfreiheit der  $\beta$ 6/ $\beta$ 7-Schleife, was dem



Hidde Ploegh erhielt seinen BSc 1975 von der Universität Groningen und promovierte 1981 an der Universität Leiden (Niederlande) über seine Arbeiten im Labor von Jack Strominger (Harvard University, USA). Ab 1980 war er Laborleiter in Köln. Nach dieser Zeit arbeitete er am Netherlands Cancer Institute (1984–1992), am MIT (1992–1997) und an der Harvard Medical School (1997–2005), bevor er zu seinem jetzigen Standort am Whitehead Institute/MIT wechselte. Er beschäftigt sich zurzeit mit dem Immunsystem und der Entwicklung chemischer Hilfsmittel zu seiner Erforschung.

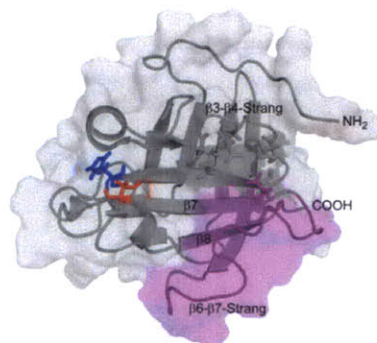


Max Popp erhielt seinen BSc in Biologie und seinen BA in Chemie von der University of Rochester (2005). Er ist zurzeit Doktorand am MIT im Labor von Hidde Ploegh, wo er daran arbeitet, die Methode der Sortase-vermittelten Proteinmodifizierung weiterzuentwickeln.



## Kurzaufsätze

H. L. Ploegh und M. W.-L. Popp



**Abbildung 2.** NMR-spektroskopisch ermittelte Struktur der Sortase A von *Staphylococcus aureus* (PDB-Code: 11JA).<sup>[16]</sup> Der aktive Cysteinrest (Cys184) ist rot und der aktive Histidinrest (His120) blau dargestellt. Die  $\beta$ 7- und  $\beta$ 8-Stränge, die die Unterseite des aktiven Zentrums bilden, sind orange, und der  $\beta$ 6- $\beta$ 7-Strang, der an der Substraterkennung beteiligt ist,<sup>[21]</sup> ist violett markiert. Calcium koordinierende Reste sind als Stabmodelle dargestellt.<sup>[27]</sup>

Substrat die Bindung ermöglicht. Die Aktivität wird so um einen Faktor 8 erhöht.<sup>[42]</sup> Zu den biochemischen Merkmalen des aktiven Zentrums zählt ein wichtiger Histidinrest (H120), der ein Thiolat-Imidazolium-Ionenpaar mit dem katalytischen Cysteinrest bildet.<sup>[43]</sup> Für die Katalyse muss die deprotonierte Form des Cysteinrestes vorliegen. Bei einem physiologischen pH-Wert befindet sich diese jedoch im Gleichgewicht mit der neutralen Form, weshalb immer nur ein kleiner Teil (ca. 0,6%) des gesamten Enzyms zu einer gegebenen Zeit wirksam sein kann.<sup>[44,45]</sup> Der Cysteinrest greift die Peptidbindung zwischen dem Threonin- und Glycinrest im Sortierungsmotiv an. Das protonierte Imidazoliumion wirkt in allgemeiner Weise als Säure für die freigesetzte Glycin- $\alpha$ -NH<sub>2</sub>-Gruppe und führt zur Bildung einer acylierten Form der Sortase. Das ankommende Glycinnukleophil wird dann deprotoniert, greift den Thioester an und bildet erneut eine Peptidbindung. Wenn stattdessen Wasser die Acylenzymzwischenstufe angreift, führt diese Reaktion irreversibel zu einem Hydrolyseprodukt.<sup>[46]</sup>

## 2. Engineering bakterieller Oberflächen

Das Sortasesystem zur Verankerung von Proteinen in der Zellwand Gram-positiver Bakterien wurde zuerst dazu verwendet, Mikroben mit heterologen Proteinen zu versehen. Solche Experimente erfordern eine genetische Fusion des heterologen Proteins mit dem Sortierungsmotiv. Das heterologe Protein wird dann exprimiert und über den normalen Zellwandsortierungsweg auf die Oberfläche gebracht. Auf diese Weise gelang die Fixierung alkalischer Phosphatase in der Zellwand von *Staphylococcus aureus*,<sup>[47]</sup> des E7-Proteins von humanem Papillomvirus 16 (HPV16) auf *Streptococcus gordonii*, einer kommensalen Mikrobe in der Mundhöhle<sup>[48]</sup> sowie der  $\alpha$ -Amylase an das Peptidoglycan von *Bacillus*

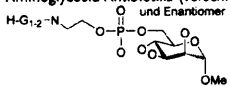
*Subtilis*, letzteres mithilfe einer Koexpression des Sortasesgens von *Listeria monocytogenes*.<sup>[49]</sup> Die Peptidoglycanzellwand kann durch Inkubation von sich teilenden *S. aureus* mit chemischen Sonden, die am N-Terminus des LPXTG-Peptids angehängt sind, sogar mit nichtbiologischen Bausteinen wie Fluorescein, Biotin oder Aziden markiert werden.<sup>[50]</sup> Beim Einbau von N-terminalen Markierungssonden (siehe Abschnitt 4) wird das endogene Sortaseenzym verwendet, das die exogen zugeführten Sonden an den Pentaglycinsseitenketten der Zellwand verankert.

## 3. C-terminale Markierung

Die Fähigkeit der Sortase zur Erkennung des Sortierungsmotivs bei dessen Einbringen in rekombinant exprimierte Proteine ermöglicht die ortsspezifische Einführung von Teilen oder funktionellen Gruppen, die genetisch nicht kodiert werden können (Abbildung 1). Diese Methode erfordert lediglich, dass das LPXTG-Motiv dem Lösungsmittel ausgesetzt ist und liefert gewöhnlich hohe Ausbeuten des gewünschten Transpeptidierungsprodukts. Mittlerweile wurden viele Proteinsubstrate mit Sonden markiert, die ein breites Spektrum an Funktionalitäten abdecken, beispielsweise Biotin, Fluorophore, Vernetzungsgagentien und multifunktionelle Sonden (Tabellen 1 und 2).<sup>[34]</sup> Das Markieren rekombinanter Proteine mit Sortase A erfordert keine hochentwickelte Syntheschemie; die meisten Sonden sind durch gewöhnliche Peptidsynthesen und mit leicht erhältlichen Reagentien zugänglich. In der Regel werden die Produktion und Faltung der rekombinanten Substratproteine nicht durch die Gegenwart der kleinen LPXTG-Markierung beeinträchtigt. Da alle Transformationen durch die Verwendung von Sortase unter physiologischen Pufferbedingungen (pH-Wert, Ionenstärke und -bedingungen) an Substraten durchgeführt werden, deren richtige Faltung und Aktivität vor der Reaktion verifiziert werden können, wird fast nie ein Verlust an biologischer Aktivität beim Endprodukt beobachtet. Die Fähigkeit, eine Sortase-katalysierte Transacylierung einzugehen, scheint allein durch die Zugänglichkeit und Flexibilität des Sortierungsmotivs bestimmt zu werden. Auf Intein beruhende Methoden des Protein-Engineerings erfordern üblicherweise, dass sich Substrate zuerst falten, während sie noch mit einer Intein-domäne fusioniert sind, die die Größe eines Proteins hat, was zu Löslichkeitsproblemen führen kann.<sup>[51]</sup>

Der Nutzen der Sortasemarkierungsmethode beruht auf der Tatsache, dass das Enzym Substrate toleriert, die in ihrer Struktur und Sequenz vor der Spaltstelle sehr unterschiedlich sein können. Diese Eigenschaft ist nicht unerwartet, wenn man bedenkt, dass Sortase fähig ist, eine große Bandbreite an Proteinsubstraten an der Zellwand zu verankern. Das Substrat muss nicht einmal ein Protein sein – Peptidnukleinsäuren (PNAs), die mit dem Sortaseerkennungsmotiv verbunden sind, können an einem Peptid, das ein Glycin trägt und zellgängig ist (amphipathisches Modellpeptid, MAP), angebracht werden, was zu aktiven, zellpenetrierenden Antisense-PNA-Peptid-Konjugaten führt.<sup>[52]</sup> Auch scheint die Identität der Substituenten in C-terminaler Position zum Glycinnukleophil

**Tabelle 1:** Beispiele für synthetische Nukleophile zur ortsspezifischen Sortase-A-Transpeptidierung.

Nr.	Sonde	Markierungsort	hinzugefügte Eigenschaft	Lit.
1	H-G <sub>3</sub> K(Biotin)-L-OH	C-terminal	biophysikalischer Griff	[32, 55]
2	H-G <sub>3</sub> K(ANP)K(Biotin)-L-OH	C-terminal	biophysikalischer Griff/Photospaltung	[32]
3	H-G <sub>3</sub> K(Phenylazid)K(Biotin)-G-OH	C-terminal	biophysikalischer Griff/Photovernetzter	[32]
4	H-G <sub>3</sub> K(FITC)-NH <sub>2</sub>	C-terminal	Fluoreszenz	[32]
5	H-G <sub>3</sub> K(K(TAMRA))-NH <sub>2</sub>	C-terminal	Fluoreszenz	[32, 71]
6	H-G <sub>3</sub> YC(Biotin)-NH <sub>2</sub>	C-terminal	biophysikalischer Griff	[33]
7	H-G <sub>3</sub> YC(Alexa 488)-NH <sub>2</sub>	C-terminal	Fluoreszenz	[33]
8	H-AA-Ahx-K(K(TAMRA))-NH <sub>2</sub>	C-terminal ( <i>S. pyogenes</i> )	Fluoreszenz	[56]
9	H-G <sub>3</sub> K(C12-C24)-NH <sub>2</sub>	C-terminal	Lipidierung	[64]
10	H-G <sub>3</sub> K(1-Ad)-NH <sub>2</sub> <sup>[a]</sup>	C-terminal	Hydrophobie	[64]
11	H-G <sub>3</sub> WK(Cholesterol)-NH <sub>2</sub>	C-terminal	Lipidierung	[64]
12	D-Tat (1. Rest ist G)	C-terminal	Zellpenetration	[53]
13	H-G <sub>2</sub> Y-PTD5-NH <sub>2</sub>	C-terminal	Zellpenetration	[53]
14	(H <sub>2</sub> NRQRRTSKLMKRAhx) <sub>2</sub> KYK(GG-NH <sub>2</sub> )-NH <sub>2</sub>	C-terminal	Zellpenetration	[53]
15	H-G <sub>3</sub> K(Folat)-NH <sub>2</sub>	C-terminal	Folsäure	[53]
16	H <sub>2</sub> N-PEG	C-terminal	inertes Polymer	[54]
17	H-G <sub>3</sub> K(PEG)-OH	C-terminal	inertes Polymer	[54]
18	H-G <sub>3</sub> -MAP-NH <sub>2</sub>	C-terminal	Zellpenetration	[52]
19	Aminoglycosid-Antibiotika (verschiedene) und Enantiomer	C-terminal	Antibiotika	[60]
20		C-terminal	GPI-Mimetikum	[61]
21	GPI-Mimetika auf Basis von 19 mit Trisaccharidkernen	C-terminal	GPI-Mimetikum	[62, 63]
22	Biotin-PEG-YGLPETGG-NH <sub>2</sub>	N-terminal	biophysikalischer Griff	[57]
23	Alexa 647-LPETGG-NH <sub>2</sub>	N-terminal	Fluoreszenz	[57]
24	Alexa 488-LPETGG-NH <sub>2</sub>	N-terminal	Fluoreszenz	[58]
25	Biotin-LPRT-OMe	N-terminal	biophysikalischer Griff	[56]
26	FITC-Ahx-LPRT-OMe	N-terminal	Fluoreszenz	[56]
27	FAM-LPETG-NH <sub>2</sub>	N-terminal	Fluoreszenz	[50]
28	Biotin-GGLPETG-NH <sub>2</sub>	N-terminal	biophysikalischer Griff	[50]
29	N <sub>3</sub> -ALPETG-NH <sub>2</sub>	N-terminal	bioorthogonale Reaktionen	[50]

[a] 1-Ad = 1-Adamantyl, Ahx = Aminohehexansäure, FAM = Carboxyfluorescein, FITC = Fluoresceinisothiocyanat, PEG = Polyethylenglycol, TAMRA = Carboxytetramethylrhodamin

ohne Bedeutung zu sein: D-Aminosäure enthaltende Peptide, Folate, verzweigte Proteintransduktionsdomänen<sup>[53]</sup> und lange Polyethylenglycolketten<sup>[54]</sup> wurden alle mithilfe von Sortase angehängt. Die Spaltstelle muss auch nicht nahe dem C-Terminus des Substratproteins sein. Eine ausreichend große Schleife, die dem Lösungsmittel zugänglich ist, genügt. Diese Eigenschaft wurde genutzt, um den Beitrag einer entscheidenden Schleife im desubiquitinierenden Enzym Ubiquitin C-Terminus Hydrolase 3 (UHL3) zur Substratbindung und Katalyse zu studieren.<sup>[55]</sup> Da die Bruchstelle in eine Schleife eingebaut werden kann, ist es möglich, das Proteinrückgrat zu durchtrennen, während gleichzeitig ein Reporter eingebaut wird (Biotin oder Fluorophor), um das Verhalten des unterbrochenen Enzyms in Gegenwart von intaktem Wildtypenzym zu untersuchen. Diese Eigenschaft trifft wahrscheinlich auf viele Proteine zu, die in einer Konformation mit flexiblen, exponierten Schleifen vorliegen.

Die C-terminale Markierung ist besonders nützlich, um Typ-II-Transmembranproteine in der lebenden Zellmembran von Säugtierzellen zu studieren. Der C-Terminus von Membranproteinen des Typs II ist dem extrazellulären Raum zugewandt und eignet sich deshalb besonders für Sortase-vermittelte Markierung. Proteine mit dieser Typ-II-Topologie sind besonders ungünstig für eine genetische Fusion mit fluoreszierenden Proteinen wie dem grün fluoreszierenden

Protein (GFP). Das Anbringen eines fluoreszierenden Proteins am N-Terminus verhindert gewöhnlich die kotranslationale Insertion von Membranproteinen des Typs II in das endoplasmatische Retikulum (ER), während die Markierung des C-Terminus mit GFP diesen sperrigen Substituenten nahe an die Stelle bringt, an der das untersuchte Typ-II-Transmembranprotein mit seinem Liganden wechselwirkt. Der CD40-Ligand (CD40 L), die Influenza-Neuraminidase<sup>[52]</sup> und der Osteoclast Differentiation Factor (ODF)<sup>[53]</sup> wurden alle auf diese Weise auf lebenden Zellen markiert.

#### 4. N-terminale Markierung

Eine Proteinmarkierung des N-Terminus gelingt ganz einfach durch Verlagerung des Sortaseerkennungselements vom Protein zur kurzen Peptidsonde und durch Einführung einer geeigneten Zahl von Glycinen am N-Terminus des Zielproteins (Abbildung 1). Sowohl Methylestermetetika des Sortasemotivs<sup>[56]</sup> als auch das komplette LPXTG-Sortaseerkennungsmotiv können als Gerüst für solche Sonden genutzt werden.<sup>[57]</sup> Diese Markierungsmethode ähnelt dem Verfahren zur Markierung des C-Terminus, außer dass das Acylenzym-Intermediat zwischen der Sortase und der Peptidsonde gebildet wird und dass das markierte Protein meh-

Tabelle 2: Beispiele für Proteine, die durch Sortase-A-Transpeptidierung markiert wurden.

Substrat	in Lösung/ an der Zelloberfläche	Markierungsart	Sonde(n) <sup>[a]</sup>	Lit.
H-2K <sup>b</sup>	Lösung	C-terminal	1,2,3,4,5	[32]
CD154	Zelloberfläche	C-terminal	1,5	[32]
Neuraminidase	Zelloberfläche	C-terminal	1	[32]
ODF	Zelloberfläche	C-terminal	6,7	[33]
Cre	Lösung	C-terminal	5	[71]
UCHL3	Lösung	C-terminal (Schleife)	1	[55]
p97	Lösung	C-terminal	5	[71]
eGFP	Lösung	C-terminal	9,10,11	[64]
GFP	Lösung	C-terminal	13,14,15	[53]
PNA	Lösung	C-terminal	18	[52]
eGFP	Lösung	C-terminal	16,17	[54]
Mrp	Lösung	C-terminal	19	[60]
YALPETGK	Lösung	C-terminal	19	[60]
(His) <sub>6</sub> YALPETGKS	Lösung	C-terminal	20	[61]
CDS2-Peptide	Lösung	C-terminal	21	[62]
CD24	Lösung	C-terminal	21	[62]
MUC1	Lösung	C-terminal	21	[63]
LPETG <sub>3</sub> -ECFP-TM	Zelloberfläche	N-terminal	22,23	[57]
LPETG <sub>3</sub> -PAFR	Zelloberfläche	N-terminal	24	[58]
G <sub>1</sub> /G <sub>2</sub> -CTXB	Lösung	N-terminal	25,26	[56]
G <sub>1</sub> -eGFP	Lösung	N-terminal	26	[56]
G-UCHL3	Lösung	N-terminal	26	[56]
<i>S. aureus</i> -Oberflächen-Peptidoglycan	Zelloberfläche	N-terminal	27,28,29	[50]
eGFP	Lösung	N- und C-terminal	26 und 8	[56]
UCHL3	Lösung	N- und C-terminal	26 und 8	[56]

[a] Sondennummer gemäß Tabelle 1.

rere Glycinreste am N-Terminus enthält, dessen  $\alpha$ -NH<sub>2</sub>-Gruppe als Nukleophil fungieren kann. Mit dieser Strategie wurde eine Fluoreszenzsonde am N-Terminus eines Membranproteins in lebenden Säugetierzellen eingeführt, nachdem zuvor das Glycinnukleophil in geschickter Weise ebenfalls mit Sortase freigesetzt worden war.<sup>[57]</sup> Dieses System wurde später dazu verwendet, Reporterfluorophore am N-Terminus des G-Protein-gekoppelten Platelet-Activating Factor Receptor (PAFR) einzubauen. Dies ermöglichte eine direkte Beobachtung des Traffickings der Rezeptoren, die sich während der Markierung auf der Zelloberfläche befanden. So wurde gezeigt, dass PAFR-Rezeptoren mit bestimmten Mutationen fehlerhaft in der Zelle wandern.<sup>[58]</sup> Die Sortasemethode zur Markierung des C- und N-Terminus von Zelloberflächenproteinen gibt lediglich Zugriff auf den Pool eines zu untersuchenden Proteins, der sich an der Zelloberfläche befindet. Dies ist ein Vorteil, wenn nur das Verhalten desjenigen Bruchteils einer Proteinmenge von Interesse ist, der der Oberfläche ausgesetzt ist. Dies ist gewöhnlich auch der relevante Teil, wenn die Bindung an Liganden auf die Zelloberfläche beschränkt ist. Proteine, die genetisch mit einem fluoreszierenden Protein fusioniert sind, sind ab dem Moment ihrer Entstehung detektierbar. Diese Eigenschaft hat zwar durchaus ihre Vorteile, kann allerdings die Unterscheidung zwischen dem Verhalten gerade synthetisierter Proteine und dem reifer, biologisch relevanter Proteine erschweren. Daher sollten die Strategien auf Sortasebasis eher als hilfreiche Ergänzung zu den auf GFP-Methoden betrachtet werden, mit dem zusätzlichen Vorteil größerer chemischer Flexibilität.

### 5. Markierung an N- und C-Terminus

Durch die Verwendung von Sortasen mit verschiedenen Substratspezifitäten ist es möglich, die N- und C-terminalen Markierungsstrategien zu kombinieren. SrtA<sub>Strept</sub> von *Streptococcus pyogenes*<sup>[59]</sup> erkennt und spaltet das LPXTA-Motiv und akzeptiert Nukleophile, die auf Alanin beruhen. Es spaltet auch – wenngleich mit geringerer Effizienz – das SrtA<sub>Staph</sub>-LPXTG-Motiv (SrtA<sub>Staph</sub>: Sortase A von *Staphylococcus aureus*). Im Unterschied dazu spaltet das SrtA<sub>Staph</sub>-Enzym das LPXTA-Motiv nicht, wodurch diese beiden Enzyme orthogonal hinsichtlich dieses Motivs sind. Diese Eigenschaft wurde genutzt, um die beiden Termini von GFP und UCHL3 mit unterschiedlichen Fluorophoren zu markieren. Hierfür wurde eine Maskierungsstrategie angewendet, in der die zur SrtA<sub>Staph</sub>-Markierung notwendigen N-terminalen Glycine durch proteolytische Spaltung mit Thrombin freigesetzt wurden. Durch diesen Schritt wurde eine Proteinoligomerisierung vermieden, die wahrscheinlich während des C-terminalen Markierungsprozesses mit SrtA<sub>Strept</sub> stattfinden würde.<sup>[56]</sup>

### 6. Imitation posttranslationaler Modifikationen

Die Sortasemethode ermöglicht die Herstellung homogener rekombinanter Proteine mit posttranslationalen Modifikationen, die nicht genetisch kodiert sind. Auf diese Weise können Glycoproteine hergestellt werden, die normalerweise durch eine Reihe komplizierter enzymatischer Prozesse im



sekretorischen Weg biosynthetisiert werden. LPXTG-markierte Proteine und Peptide können mit Nucleophilen modifiziert werden, die auf 6-Amino-hexose beruhen, darunter Aminoglycosid-abgeleitete Antibiotika und ihre Analoga.<sup>[60]</sup> Glycosylphosphatidylinositol(GPI)-Anker, die normalerweise am C-Terminus des Proteins angebracht sind, können durch Bindung von LPXTG-Peptiden an synthetische Glycinnucleophile, die ihrerseits an die Phosphoethanolamineinheit eines GPI-Derivats gebunden sind, imitiert werden.<sup>[61]</sup> Eine Reihe von Peptiden (CD52-Fragment, Mucin-1-Peptide) und kleinen Proteinen (CD24) wurde so an GPI-Mimetika mit Trisaccharidkernen angehängt.<sup>[62,63]</sup> Lipidmodifikationen von Proteinen sind weitere wichtige posttranslationale Veränderungen, die wenig studiert sind, da es an Methoden mangelt, um homogene Präparationen von Lipoproteinen zu erhalten. Zur Lösung dieses Problems wurde Sortase verwendet.<sup>[64]</sup> Ein Gerüst auf Glycinbasis wurde mit einer Auswahl an linearen Alkylketten (C<sub>12</sub>-C<sub>24</sub>), Cholesterin sowie Adamantan modifiziert und anschließend dazu verwendet, geeignete Versionen von enhanced GFP (eGFP) zu modifizieren, die mit LPETG markiert waren. Diese eGFP-Lipoproteine assoziierten abhängig von der Kettenlänge mit den Plasmamembranen lebender Zellen (wobei eine C<sub>22</sub>-Kette das beste Resultat lieferte), von wo aus sie Zugang zum endosomalen Kompartiment fanden.

### 7. Schrittweiser Aufbau von Proteinen, Proteindomänen und Peptiden

Gefaltete Proteine mit einem exponierten Glycinrest am N-Terminus können als Nucleophile für eine Markierung mit Sortase verwendet werden. Substratproteine mit LPXTG-Motiv können mit dem ankommenden Nucleophilprotein fusioniert werden, was zu großen Transpeptidierungsprodukten führt. Durch Verwendung unabhängig gefalteter Proteine als Substrate ist es möglich, viele der Löslichkeitsprobleme zu vermeiden, die die Expression genetisch kodierter Fusionen erschweren. Diese Eigenschaft wurde genutzt, um die Strukturanalyse mit NMR-Spektroskopie zu erleichtern, bei der geringe Löslichkeit ein Hauptproblem ist, da hier gewöhnlich hohe Proteinkonzentrationen notwendig sind. Mithilfe von Sortase wurde eine unmarkierte und daher für die NMR-Spektroskopie unsichtbare, löslichkeitserhöhende Domäne auf Basis der B1-Immunglobulin-Bindungsdomäne (G<sub>3</sub>-GB1) an den C-Terminus der Vav-SRC-Homologie-3-Domäne (SH3; eine Domäne, die um pH 7 nahezu unlöslich ist) angebracht.<sup>[65]</sup> Die Struktur der <sup>13</sup>C/<sup>15</sup>N-markierten, angehefteten Vav-SH3-Domäne wurde im Anschluss NMR-spektroskopisch ohne Störsignale vonseiten der löslichkeitserhöhenden Gruppe aufgeklärt. Auch über die teilweise Markierung des MecA-Proteins von *Bacillus subtilis* durch Verwendung von Sortase in einer NMR-spektroskopischen Studie wurde berichtet.<sup>[66]</sup> Mit Sortase wurden auch zahlreiche Immundetektionsreagentien erzeugt. Weiterhin gelangen Protein-Protein-Ligationen zwischen dem Fc-Bindungsmodul (ZZ-Domäne) und mehreren Detektionsenzymen (alkalische Phosphatase, Luciferase, Glucoseoxidase) mit Sortase.<sup>[67]</sup> Außerdem wurden Mucin-Glycopeptide syn-

thetisiert, die N- und O-gebundene Glycane enthalten. Durch eine Kombination von chemischer Synthese mit Glycanstrukturaufklärung durch enzymatische Synthese wurden unterschiedliche Peptide aufgebaut, die N- und O-gebundene Glycane enthalten. Diese Glycopeptide wurden anschließend durch Sortase zu einem stereochemisch homogenen Präparat verbunden.<sup>[68,69]</sup> Sortase wurde in einer kombinierten Strategie aus rekombinanter, enzymatischer und chemischer Synthese (CRECS) dazu verwendet, Mimetika von G-Proteingekoppelten Rezeptoren (GPCR) herzustellen.<sup>[70]</sup> Die GPCRs sind Proteine mit sieben Transmembrandomänen, die ihre Liganden über drei extrazelluläre Schleifen und das extrazelluläre N-terminale Segment binden. Zur Nachahmung dieser Anordnung wurden drei Schleifen synthetisiert, die durch native chemische Ligation cyclisiert und an ein Peptidgerüst mit einem Triglycinmotiv angehängt wurden. Im Anschluss wurde der N-Terminus des GPCR, der die Sortasespaltsequenz trug, in *E. coli* exprimiert und durch eine Sortase-vermittelte Ligation an das Gerüst angehängt. Diese elegant entworfenen, löslichen Mimetika sollten eine systematische Charakterisierung der Beiträge jeder Region zur Ligandenbindung ermöglichen und demonstrieren die Möglichkeiten einer Kombination von chemischer Synthese mit Molekularbiologie.

Eine einzigartige Variante der Protein-Protein-Ligation findet statt, wenn das LPXTG-Motiv und N-terminale Glycine im selben Konstrukt vorhanden sind. Wenn die beiden Einheiten im gefalteten Protein einander nahe genug kommen, kann ein zirkuläres Transpeptidierungsprodukt gebildet werden, indem der N-Terminus mit dem Sortaseerkenntnis-element eine stabile Bindung eingeht.<sup>[54,71]</sup> Zirkuläre Proteine haben nützliche biochemische Eigenschaften: Sie sind widerstandsfähig gegen Aggregation, benötigen mehr Energie für die Denaturierung und sind, da sie keinen exponierten Terminus haben, in ihrem natürlichen Zustand vor einem Exoproteasenangriff geschützt.<sup>[72-76]</sup>

### 8. Verankerung an festen Oberflächen

Auch eine kovalente Immobilisierung an festen Oberflächen wurde durch Sortase erreicht. Ein Hauptvorteil dieser Methode ist, dass dank der Spezifität des Sortaseenzym Proteine einheitlich und mit definierter Orientierung für die nachfolgende Analyse an festen Oberflächen verankert werden können. Wegen der starken Amidbindung zwischen Oberfläche und Protein können auch harsche Waschbedingungen angewendet werden, z. B. für die kovalente Bindung von GFP an Glycin-derivatisierte Polystyrolkugeln.<sup>[54]</sup> Auf ähnliche Weise wurden Adhäsionsproteine von Gram-positiven Bakterien an fluoreszierende, Glycin-derivatisierte Polystyrolkugeln gebunden.<sup>[77]</sup> Ebenfalls gelang eine Verankerung von GFP und Tus-Proteinen an Glycidylmethacrylatkugeln, die mit Oligoglycin derivatisiert waren, sowie an Glycin-modifizierter Agarose (Affigel) und an Glycin-modifiziertem Aminosilan auf Glasplättchen.<sup>[78]</sup> Ein weiteres Beispiel ist die orientierte Verankerung von Proteinen an Triglycin-modifizierten, carboxymethylierten Dextran-Biacore-Sensorchips für die Oberflächenplasmonenresonanz-

## Kurzaufsätze

H. L. Ploegh und M. W.-L. Popp

spektroskopie.<sup>[79]</sup> Rekombinantes Fibronectin-Bindungsprotein (rFba-LPETG) von A-Streptokokken (Group A Streptococcus, GAS) wurde auf diese Art verankert, was die Messung der Bindung des humanen Faktors H an immobilisiertes Protein ermöglichte. Für die Entwicklung von Reagentien zur chemoenzymatischen Synthese von Glycokonjugaten wurden immobilisierte  $\beta$ -1,4-Galactosyltransferase (rhGalT) und  $\alpha$ -1,3-Fucosyltransferase (rHFucT) aus *Helicobacter pylori* kovalent an Alkylamin-Sepharosekügelchen befestigt. Diese Enzyme sind aktiv und wiederverwendbar, wenn sie mit definierter Ausrichtung an der festen Oberfläche verankert sind.<sup>[80]</sup>

## 9. Proteinexpression und -aufreinigung

Genetische Fusionen von Sortase mit anderen Proteinen wurden zum Zwecke der Proteinaufreinigung durchgeführt. Ein hochreines Protein mit einem zusätzlichen Glycinrest<sup>[81]</sup> wurde erhalten, indem Hexahistidin-markierte Sortase, gefolgt von LPETG, mit dem zu untersuchenden Protein linear fusioniert wurde, durch Metallchelate-Affinitätschromatographie an einer Nickelnitrilotriacetatmatrix (Ni-NTA-IMAC) aufgereinigt wurde und zum Schluss durch Versetzen mit  $\text{Ca}^{2+}$  und Triglycin von der festen Phase abgespalten wurde (Abbildung 3). Mit dem Ziel einer allgemeinen Reinigungsme-



**Abbildung 3.** Proteinaufreinigung mit Sortase A. Rekombinant exprimierte Proteine werden als Fusionsproteine hergestellt, die entweder eine Hexahistidinmarkierung<sup>[81]</sup> (His<sub>6</sub>, oben) oder ein Biotinakzeptorpeptid<sup>[82]</sup> (AP, unten) enthalten, gefolgt vom katalytischen Zentrum der Sortase, der LPXTG-Markierung und dem zu untersuchenden Protein. Versetzen des immobilisierten Fusionsproteins mit  $\text{Ca}^{2+}$ -Ionen und Oligoglycin stimuliert die Sortaseaktivität. Das zu untersuchende Protein wird als gereinigtes Produkt mit einem zusätzlichen N-terminalen Glycinrest freigesetzt.

thode für die großtechnische Proteinherstellung wurde diese Methode an ein zellfreies Weizenkeim-Translationssystem angepasst. Bei dieser Version wird die Hexahistidinmarkierung durch ein Biotinakzeptorpeptid ersetzt (ein 15 Aminosäuren langes Peptid, das von der *E. coli*-Biotinligase erkannt wird), und die Proteine werden mit immobilisiertem Streptavidin in Gegenwart von Calciumchelatoren gereinigt, um eine vorzeitige Abspaltung zu vermeiden.<sup>[82]</sup> Beide Methoden führen zu einem Zielprotein mit einem zusätzlichen Glycinrest am N-Terminus, eine Konfiguration, die für die N-terminale Markierung durch Sortase sehr geeignet ist (Abbildung 3).

Sortase A von *S. aureus* ist ein sehr gut lösliches Enzym, das in hoher Ausbeute ( $>40 \text{ mg L}^{-1}$  Kultur) hergestellt werden kann. Diese Eigenschaft wurde genutzt, um die Löslichkeit von Proteinen durch Fusion mit einer Sortaseversion ohne katalytischen Cysteinrest zu erhöhen.<sup>[83]</sup>

## 10. Ausblick

In der mikrobiologischen Welt gibt es eine Vielfalt von Sortasen mit unterschiedlichen Präferenzen für Spaltsequenz und Nukleophil.<sup>[84]</sup> Zum Beispiel akzeptieren die Pilin aufbauenden Sortasen die  $\epsilon$ -Aminogruppe eines Lysins in der YPKN-Sequenz als Nukleophil, die Sortase A von *S. Aureus* dagegen den N-Terminus von Glycinanhängen. Sortase B (SrtB) von *S. aureus* spaltet die NPQTN-Sequenz.<sup>[85,86]</sup> während SrtB von *Bacillus anthracis* die NPKTG-Sequenz spaltet<sup>[87]</sup> und meso-Diaminopimelinsäure als Nukleophil akzeptiert.<sup>[88-90]</sup> Noch befindet sich der Einsatz von Sortasen für die Proteinmarkierung in seinen Anfängen, allerdings sind die Perspektiven vielversprechend, wenn die erforderlichen Reaktionsbedingungen und die Bedingungen für die Substraterkennung erst einmal besser geklärt sind. Mit dem Aufkommen zahlreicher Techniken zur gesteuerten Proteinevolution wird sich vielleicht verbessern lassen, was uns die Natur zur Verfügung gestellt hat. In ihrem gewöhnlichen biologischen Zusammenhang verankert die Housekeeping-Sortase strukturell sehr unterschiedliche Substrate in der Zellwand.<sup>[91]</sup> Die einzige Anforderung für den Sortaseangriff scheint das Vorhandensein des LPXTG-Pentapeptids zu sein. Dieser Befund – zusammen mit der Tatsache, dass die Substraterkennung primär eine Funktion der  $\beta$ -6-7-Schleife ist<sup>[92-94]</sup> – lässt vermuten, dass sich möglicherweise Versionen von Sortase entwickeln lassen, die Peptidsequenzmotive komplett orthogonal angreifen können. Wenn man die Vielfalt an Sortasen in Betracht zieht, die verschiedene Ankerstrukturen erkennen können,<sup>[88,89]</sup> sollte die Entwicklung von Enzymen möglich sein, die nichtnatürliche Nukleophile akzeptieren. Vorstellbar ist auch eine orthogonale Markierung verschiedener Proteine in einem Reaktionsansatz oder ihrer biologischen Äquivalente in derselben Zellumgebung durch Kombination von Sortasen unterschiedlicher Spezifität. Die Geschwindigkeit der Sortasereaktion mit SrtA von *S. aureus* ist bestenfalls mäßig, wahrscheinlich weil das Gleichgewicht des aktiven Zentrums auf der Seite des nichtaktiven, nichtionisierten Cysteinrestes liegt. Zwar kann dieses Problem leicht durch Verwendung großer Mengen an Sortase (und dadurch einer größeren Gesamtmenge des katalysierenden Enzyms) überwunden werden, jedoch wäre es von Vorteil, die Gesamtkinetik durch gesteuerte Proteinevolution zu verbessern. Die Verwendung der Pilus bildenden Sortasen, die Isopeptidbindungen durch Polymerisation von Untereinheiten aufbauen, sollte noch eingehender erforscht werden. Diese Enzyme könnten von besonderem Interesse sein, da sie sich entwickelt haben, um posttranslational kovalente Protein-Protein-Bindungen zu bilden. Diese Eigenschaft könnte dazu genutzt werden, um in vitro ähnliche Protein-Protein-Ligationen durchzuführen und interne, exponierte Lysinreste als ankommende Nukleophile zu nutzen. Da Isopeptidbindun-

8

www.angewandte.de

© 2011 Wiley-VCH Verlag GmbH &amp; Co. KGaA, Weinheim

Angew. Chem. 2011, 123, 2–12

These are not the final page numbers!

gen in vielen biologischen Prozessen vorkommen, z. B. im Ubiquitinsystem, wäre die Möglichkeit zur Bildung homogener ortsspezifischer Isopeptidbindungen sehr nützlich.

Ein interessante Frage ist, wie man die Hindernisse für die intrazelluläre Markierung durch Sortase überwinden könnte. Es ist möglich, die Sortase innerhalb lebender bakterieller und eukaryotischer Zellen zu exprimieren – das Enzym faltet sich richtig und ist aktiv.<sup>[95]</sup> Das *S. pyogenes*-Enzym, das keine zweiwertigen Kationen benötigt, ist in Calcium-freien Kompartimenten aktiv, z. B. im Cytosol. Ähnlich dazu können Substrate, die ein exponiertes LPXTG-Motiv aufweisen, leicht exprimiert werden und – abhängig vom Protein – die Lokalisierung und/oder Funktion ihrer unmodifizierten Variante beibehalten. Wenn das Nukleophil für die Transpeptidierung genetisch als Teil eines Substratproteins kodiert ist, wie im Fall von Zirkularisierungsreaktionen oder Protein-Protein-Ligationsreaktionen, verläuft die Transpeptidierung leicht<sup>[95]</sup> und führt in Hefezellen, Säugtierzellen und Prokaryoten intrazellulär zu zirkularisierten Proteinen. Der geschwindigkeitsbestimmende Schritt für die einfache Markierung von C-Termini ist deshalb die Verfügbarkeit eines exogenen Nukleophils in Mengen, die hoch genug sind, um das Transpeptidierungsprodukt biochemisch analysieren zu können. Diese Begrenzung könnte dadurch überwunden werden, dass man eine der vielen Proteintransduktionssequenzen des HIV-TAT-Proteins (TAT: Trans-Activator of Transcription) oder sonstiger Quellen am markierten Nukleophil befestigt. Diese Möglichkeit vernachlässigt jedoch den weiteren Verbleib nichtinkorporierter cytoplasmischer Nukleophile und wurde bisher nicht in die Praxis umgesetzt. Andererseits wäre es vielleicht möglich, die Sonden irreversibel zuzuführen, indem man die  $\alpha$ -Aminogruppe und andere Gruppen, die die Translokation ins Cytosol verhindern, mithilfe von chemischen Gruppen temporär maskiert, indem man chemische Gruppierungen einführt, die sich durch Esterasen, photochemisch oder im reduzierenden Milieu des Cytoplasmas wieder abspalten lassen.

Ein wesentlicher Vorteil der Markierung mit Sortase ist, dass die nukleophilen Sonden relativ leicht synthetisiert werden können. Die meisten Sonden sind einfache, C-terminale Glycinanhänge, die durch chemische Routinekupplungen an einer festen Phase oder in Lösung hergestellt werden. Literaturdaten zeigen, dass viele derartige Sonden wahrscheinlich in gleicher Weise als Nukleophile mit einer gegebenen Zielstruktur reagieren. Dies macht es möglich, eine ganze Reihe von Biokonjugaten herzustellen, ohne die Notwendigkeit einer Neuoptimierung jedes einzelnen Konjugats. Transportproteine, mit denen sich Bibliotheken von small interfering RNA (siRNA) und Antikörper-Medikament-Konjugaten zell- und gewebespezifisch verabreichen lassen, könnten in Reichweite sein. Ein zusätzlicher Vorteil der Sortasemethode liegt in der genauen Steuerbarkeit der Stöchiometrie des Markierungsprozesses, was quantitative Arbeiten zu solchen Biokonjugatbibliotheken sehr viel einfacher machen sollte.

Außerdem ermöglicht es Sortase auch, die Orientierung von Ligationspartnern zu steuern, was bei der Herstellung von Protein-Protein-Konjugaten nützlich sein könnte. Durch Austausch der N-terminalen Glycine einer Einheit gegen das

Sortasemotiv des Partnerproteins kann die Orientierung der Untereinheiten umgekehrt werden. Durch klug entworfene Sonden, die „Griffe“ für bioorthogonale Reaktionen enthalten (z. B. die Reaktion eines Azids mit einem Alkin, eines Azids mit einem gespannten Cyclooctin oder eines Aldehyds mit einer Aminoxygruppe), sollte es möglich sein, Proteine aufzubauen, die kovalent in nichtnatürlicher Topologie (N-N- oder C-C-Verknüpfung) verbunden sind. Alternativ könnte man zweistufige Transacylierungsprozesse mit den semi-orthogonalen SrtA<sub>Staph</sub>- und SrtA<sub>Strep</sub>-Enzymen sowie passenden Peptidsonden für N-N- oder C-C-Bindungen nutzen. Die Sortasemethode eröffnet damit den Zugang zu Proteinbindungen, die nicht in der Natur vorkommen. Ein Gebiet, das von dieser Möglichkeit profitieren kann, ist die Konstruktion spezieller Liganden zum Studium der Signaltransduktion in Zellen. Es wäre möglich, chimäre, gefaltete Proteinliganden zu entwerfen, die die Bindung und Heterodimerisierung ihrer zugehörigen Rezeptoren einschränken. Die leichte Anwendbarkeit und die Flexibilität der Sortasemarkierungsmethode werden mit großer Sicherheit noch weitere hochinteressante neue Forschungsbereiche erschließen.

Eingegangen am 30. Dezember 2010

Online veröffentlicht am ■■ ■■ ■■■■

Übersetzt von Dr. Gerhard Popp und Paul-Albert König

- [1] L. A. Marraffini, A. C. Dedent, O. Schneewind, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 192.
- [2] S. K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, *Science* **1999**, *285*, 760.
- [3] S. K. Mazmanian, H. Ton-That, O. Schneewind, *Mol. Microbiol.* **2001**, *40*, 1049.
- [4] S. D. Zink, D. L. Burns, *Infect. Immun.* **2005**, *73*, 5222.
- [5] H. Ton-That, O. Schneewind, *J. Biol. Chem.* **1999**, *274*, 24316.
- [6] H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12424.
- [7] R. G. Kruger, B. Otvos, B. A. Frankel, M. Bentley, P. Dostal, D. G. McCafferty, *Biochemistry* **2004**, *43*, 1541.
- [8] S. Dramsi, P. Trieu-Cuot, H. Bierne, *Res. Microbiol.* **2005**, *156*, 289.
- [9] M. G. Pucciarelli, E. Calvo, C. Sabet, H. Bierne, P. Cossart, F. Garcia-del Portillo, *Proteomics* **2005**, *5*, 4808.
- [10] D. Comfort, R. T. Clubb, *Infect. Immun.* **2004**, *72*, 2710.
- [11] a) Internet-Adresse: [http://bamics3.cmbi.kun.nl/cgi-bin/jos/sortase\\_substrates/index.py](http://bamics3.cmbi.kun.nl/cgi-bin/jos/sortase_substrates/index.py); b) J. Boekhorst, M. W. de Been, M. Kleerebezem, R. J. Siezen, *J. Bacteriol.* **2005**, *187*, 4928.
- [12] M. J. Pallen, A. C. Lam, M. Antonio, K. Dunbar, *Trends Microbiol.* **2001**, *9*, 97.
- [13] H. Ton-That, S. K. Mazmanian, K. F. Faull, O. Schneewind, *J. Biol. Chem.* **2000**, *275*, 9876.
- [14] A. Aulabaugh, W. Ding, B. Kapoor, K. Tabei, L. Alksne, R. Dushin, T. Zatz, G. Ellestad, X. Huang, *Anal. Biochem.* **2007**, *360*, 14.
- [15] A. M. Perry, H. Ton-That, S. K. Mazmanian, O. Schneewind, *J. Biol. Chem.* **2002**, *277*, 16241.
- [16] H. Ton-That, O. Schneewind, *Trends Microbiol.* **2004**, *12*, 228.
- [17] J. R. Scott, D. Zahner, *Mol. Microbiol.* **2006**, *62*, 320.
- [18] A. Mandlik, A. Swierczynski, A. Das, H. Ton-That, *Trends Microbiol.* **2008**, *16*, 33.
- [19] T. Proft, E. N. Baker, *Cell. Mol. Life Sci.* **2009**, *66*, 613.
- [20] H. Ton-That, O. Schneewind, *Mol. Microbiol.* **2003**, *50*, 1429.
- [21] H. Ton-That, L. A. Marraffini, O. Schneewind, *Mol. Microbiol.* **2004**, *53*, 251.

- [22] C. Manzano, C. Contreras-Martel, L. El Mortaji, T. Izore, D. Fenel, T. Vernet, G. Schoehn, A. M. Di Guilmi, A. Dessen, *Structure* **2008**, *16*, 1838.
- [23] F. Neiers, C. Madhurantakam, S. Falcker, C. Manzano, A. Dessen, S. Normark, B. Henriques-Normark, A. Achour, *J. Mol. Biol.* **2009**, *393*, 704.
- [24] C. Manzano, T. Izore, V. Job, A. M. Di Guilmi, A. Dessen, *Biochemistry* **2009**, *48*, 10549.
- [25] J. M. Budzik, L. A. Marraffini, P. Souda, J. P. Whitelegge, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10215.
- [26] A. Mandlik, A. Das, H. Ton-That, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14147.
- [27] S. K. Mazmanian, G. Liu, E. R. Jensen, E. Lenoy, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5510.
- [28] G. K. Paterson, T. J. Mitchell, *Trends Microbiol.* **2004**, *12*, 89.
- [29] A. W. Maresso, O. Schneewind, *Pharmacol. Rev.* **2008**, *60*, 128.
- [30] N. Suree, M. E. Jung, R. T. Clubb, *Mini-Rev. Med. Chem.* **2007**, *7*, 991.
- [31] G. K. Paterson, T. J. Mitchell, *Microbes Infect.* **2006**, *8*, 145.
- [32] M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, *Nat. Chem. Biol.* **2007**, *3*, 707.
- [33] T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, *ChemBioChem* **2008**, *9*, 802.
- [34] M. W. Popp, J. M. Antos, H. L. Ploegh, *Curr. Protoc. Protein Sci.* **2009**, Kap. 15, Abschnitt 153.
- [35] U. Ilangovan, J. Iwahara, H. Ton-That, O. Schneewind, R. T. Clubb, *J. Biomol. NMR* **2001**, *19*, 379.
- [36] U. Ilangovan, H. Ton-That, J. Iwahara, O. Schneewind, R. T. Clubb, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6056.
- [37] Y. Zong, T. W. Bice, H. Ton-That, O. Schneewind, S. V. Narayana, *J. Biol. Chem.* **2004**, *279*, 31383.
- [38] N. Suree, C. K. Liew, V. A. Villareal, W. Thieu, E. A. Fadeev, J. J. Clemens, M. E. Jung, R. T. Clubb, *J. Biol. Chem.* **2009**, *284*, 24465.
- [39] R. G. Kruger, P. Dostal, D. G. McCafferty, *Anal. Biochem.* **2004**, *326*, 42.
- [40] L. A. Marraffini, H. Ton-That, Y. Zong, S. V. Narayana, O. Schneewind, *J. Biol. Chem.* **2004**, *279*, 37763.
- [41] B. A. Frankel, Y. Tong, M. L. Bentley, M. C. Fitzgerald, D. G. McCafferty, *Biochemistry* **2007**, *46*, 7269.
- [42] M. T. Naik, N. Suree, U. Ilangovan, C. K. Liew, W. Thieu, D. O. Campbell, J. J. Clemens, M. E. Jung, R. T. Clubb, *J. Biol. Chem.* **2006**, *281*, 1817.
- [43] H. Ton-That, S. K. Mazmanian, L. Alksne, O. Schneewind, *J. Biol. Chem.* **2002**, *277*, 7447.
- [44] K. M. Connolly, B. T. Smith, R. Pilpa, U. Ilangovan, M. E. Jung, R. T. Clubb, *J. Biol. Chem.* **2003**, *278*, 34061.
- [45] B. A. Frankel, R. G. Kruger, D. E. Robinson, N. L. Kelleher, D. G. McCafferty, *Biochemistry* **2005**, *44*, 11188.
- [46] X. Huang, A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei, G. Ellestad, *Biochemistry* **2003**, *42*, 11307.
- [47] O. Schneewind, P. Model, V. A. Fischetti, *Cell* **1992**, *70*, 267.
- [48] G. Pozzi, M. Contorni, M. R. Oggioni, R. Manganelli, M. Tommasino, F. Cavalieri, V. A. Fischetti, *Infect. Immun.* **1992**, *60*, 1902.
- [49] H. D. Nguyen, W. Schumann, *J. Biotechnol.* **2006**, *122*, 473.
- [50] J. W. Nelson, A. G. Chamessian, P. J. McEnaney, R. P. Murelli, B. I. Kazmierczak, D. A. Spiegel, *ACS Chem. Biol.* **2010**, *5*, 1147.
- [51] M. Vila-Perelló, T. W. Muir, *Cell* **2010**, *143*, 191.
- [52] S. Pritz, Y. Wolf, O. Kraetke, J. Klose, M. Bienert, M. Beyersmann, *J. Org. Chem.* **2007**, *72*, 3909.
- [53] H. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670.
- [54] R. Parthasarathy, S. Subramanian, E. T. Bodcr, *Bioconjugate Chem.* **2007**, *18*, 469.
- [55] M. W. Popp, K. Artavanis-Tsakonas, H. L. Ploegh, *J. Biol. Chem.* **2009**, *284*, 3593.
- [56] J. M. Antos, G. L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W. Popp, H. L. Ploegh, *J. Am. Chem. Soc.* **2009**, *131*, 10800.
- [57] T. Yamamoto, T. Nagamune, *Chem. Commun.* **2009**, 1022.
- [58] N. Hirota, D. Yasuda, T. Hashidate, T. Yamamoto, S. Yamaguchi, T. Nagamune, T. Nagase, T. Shimizu, M. Nakamura, *J. Biol. Chem.* **2010**, *285*, 5931.
- [59] P. R. Race, M. L. Bentley, J. A. Melvin, A. Crow, R. K. Hughes, W. D. Smith, R. B. Sessions, M. A. Kchoec, D. G. McCafferty, M. J. Banfield, *J. Biol. Chem.* **2009**, *284*, 6924.
- [60] S. Samantaray, U. Marathe, S. Dasgupta, V. K. Nandicoori, R. P. Roy, *J. Am. Chem. Soc.* **2008**, *130*, 2132.
- [61] X. Guo, Q. Wang, B. M. Swartz, Z. Guo, *J. Am. Chem. Soc.* **2009**, *131*, 9878.
- [62] Z. Wu, X. Guo, Q. Wang, B. M. Swartz, Z. Guo, *J. Am. Chem. Soc.* **2010**, *132*, 1567.
- [63] Z. Wu, X. Guo, Z. Guo, *Chem. Commun.* **2010**, *46*, 5773.
- [64] J. M. Antos, G. M. Miller, G. M. Grotenbreg, H. L. Ploegh, *J. Am. Chem. Soc.* **2008**, *130*, 16338.
- [65] Y. Kobashigawa, H. Kumeta, K. Ogura, F. Inagaki, *J. Biomol. NMR* **2009**, *43*, 145.
- [66] M. A. Refaei, A. Combs, D. J. Kojetin, J. Cavanagh, C. Caparelli, M. Rance, J. Sapiro, P. Tsang, *J. Biomol. NMR* **2011**, *49*, 3.
- [67] T. Sakamoto, S. Sawamoto, T. Tanaka, H. Fukuda, A. Kondo, *Bioconjugate Chem.* **2010**, *21*, 2227.
- [68] T. Matsushita, S. Nishimura, *Methods Enzymol.* **2010**, *478*, 485.
- [69] T. Matsushita, R. Sadamoto, N. Ohyabu, H. Nakata, M. Fumoto, N. Fujitani, Y. Takegawa, T. Sakamoto, M. Kuroguchi, H. Hinou, H. Shimizu, T. Ito, K. Naruchi, H. Togame, H. Takemoto, H. Kondo, S. Nishimura, *Biochemistry* **2009**, *48*, 11117.
- [70] S. Pritz, O. Kraetke, A. Klose, J. Klose, S. Rothmund, K. Fehner, M. Bienert, M. Beyersmann, *Angew. Chem.* **2008**, *120*, 3698; *Angew. Chem. Int. Ed.* **2008**, *47*, 3642.
- [71] J. M. Antos, M. W. Popp, R. Ernst, G. L. Chew, E. Spooner, H. L. Ploegh, *J. Biol. Chem.* **2009**, *284*, 16028.
- [72] R. J. Clark, H. Fischer, L. Dempster, N. L. Daly, K. J. Rosengren, S. T. Nevin, F. A. Meunier, D. J. Adams, D. J. Craik, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13767.
- [73] R. J. Clark, J. Jensen, S. T. Nevin, B. P. Callaghan, D. J. Adams, D. J. Craik, *Angew. Chem.* **2010**, *122*, 6695; *Angew. Chem. Int. Ed.* **2010**, *49*, 6545.
- [74] M. Trabi, D. J. Craik, *Trends Biochem. Sci.* **2002**, *27*, 132.
- [75] D. J. Craik, *Science* **2006**, *311*, 1563.
- [76] A. S. Andersen, E. Palmqvist, S. Bang, A. C. Shaw, F. Hubalek, U. Ribel, T. Hoeg-Jensen, *J. Pept. Sci.* **2010**, *16*, 473.
- [77] S. Wu, T. Proft, *Biotechnol. Lett.* **2010**, *32*, 1713.
- [78] L. Chan, H. F. Cross, J. K. She, G. Cavalli, H. F. Martins, C. Neylon, *PLoS One* **2007**, *2*, e1164.
- [79] F. Clow, J. D. Fraser, T. Proft, *Biotechnol. Lett.* **2008**, *30*, 1603.
- [80] T. Ito, R. Sadamoto, K. Naruchi, H. Togame, H. Takemoto, H. Kondo, S. Nishimura, *Biochemistry* **2010**, *49*, 2604.
- [81] H. Mao, *Protein Expression Purif.* **2004**, *37*, 253.
- [82] S. Matsunaga, K. Matsuoka, K. Shimizu, Y. Endo, T. Sawasaki, *BMC Biotechnol.* **2010**, *10*, 42.
- [83] J. Caswell, P. Snoddy, D. McMeel, R. J. Buick, C. J. Scott, *Protein Expression Purif.* **2010**, *70*, 143.
- [84] T. C. Barnett, A. R. Patel, J. R. Scott, *J. Bacteriol.* **2004**, *186*, 5865.
- [85] L. A. Marraffini, O. Schneewind, *J. Biol. Chem.* **2005**, *280*, 16263.
- [86] S. K. Mazmanian, H. Ton-That, K. Su, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2293.
- [87] A. W. Maresso, T. J. Chapa, O. Schneewind, *J. Bacteriol.* **2006**, *188*, 8145.
- [88] A. H. Gaspar, L. A. Marraffini, E. M. Glass, K. L. Debord, H. Ton-That, O. Schneewind, *J. Bacteriol.* **2005**, *187*, 4646.

## Chapter 8: Future Directions

Protein-Engineering mit Sortasen

**Angewandte**  
Chemie

- [89] J. M. Budzik, S. Y. Oh, O. Schneewind, *J. Biol. Chem.* **2008**, 283, 36676.
- [90] E. M. Weiner, S. Robson, M. Marohn, R. T. Clubb, *J. Biol. Chem.* **2010**, 285, 23433.
- [91] R. Janulczyk, M. Rasmussen, *Infect. Immun.* **2001**, 69, 4019.
- [92] C. K. Liew, B. T. Smith, R. Pilpa, N. Suree, U. Ilangovan, K. M. Connolly, M. E. Jung, R. T. Clubb, *FEBS Lett.* **2004**, 571, 221.
- [93] M. L. Bentley, H. Gaweska, J. M. Kielec, D. G. McCafferty, *J. Biol. Chem.* **2007**, 282, 6571.
- [94] M. L. Bentley, E. C. Lamb, D. G. McCafferty, *J. Biol. Chem.* **2008**, 283, 14762.
- [95] K. Strijbis, M. W. Popp, unveröffentlichte Ergebnisse.
-

## Kurzaufsätze

H. L. Ploegh und M. W.-L. Popp

### Kurzaufsätze

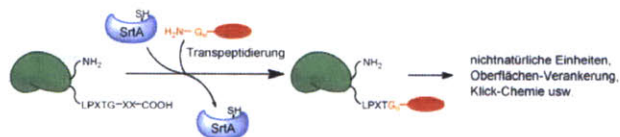
Protein-Engineering

M. W.-L. Popp,

H. L. Ploegh\*



Bilden und Brechen von  
Peptidbindungen: Protein-Engineering  
mithilfe von Sortase



**Peptide bauen mit Sortase:** Die enzymatische Bildung von Peptidbindungen mit der Transpeptidase Sortase A (SrtA) ist ein praktisches und mildes Verfahren, um Proteine so zu manipulieren, dass sie Modifikationen enthalten, die nicht genetisch kodiert sind. Dies ermöglicht

zahlreiche Anwendungen, von der homogenen Herstellung imitierter posttranslationaler Modifikationen über das Zusammenbauen von Proteindomänen bis zur Verankerung von Proteinen an festen Oberflächen.

nichtnatürliche Einheiten,  
Oberflächen-Verankerung,  
Klick-Chemie usw.

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

**Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages**

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

**Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages**

(Unpublished, from: Alexandre Esteban, Maximilian W. Popp, Karin Strijbis, Hidde L. Ploegh, and Gerald R. Fink; Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages, *submitted*)

**Abstract**

Dectin-1, the major  $\beta$ -glucan receptor in leukocytes, triggers an effective immune response upon fungal recognition. Although dectin-1 plays a crucial role in the fight against fungal infections, most aspects of its signaling *in vivo* still remain obscure. Here we show that murine dectin-1 can be labeled on live cells through sortase-mediated transpeptidation, a procedure that places a small peptide tag on the protein. Tagged dectin-1 containing a sortase A recognition domain transits to the membrane and is functional. Installation of probes by sortagging permitted highly specific visualization of functional dectin-1 on the cell surface and internalized upon presentation of a ligand. Sortagging made it possible to uncover a novel interaction between dectin-1 and galectin-3 in macrophages that increases in response to a fungal challenge. This association may be key in the modulation of the proinflammatory response toward pathogenic fungi. Our findings further strengthen the concept that pattern recognition receptors often engage multiple binding partners to generate the desired functional outcome.

**Introduction**

Dectin-1, the primary receptor on leucocytes for phagocytosis of many diverse fungi<sup>1-2</sup>, recognizes diverse fungal species including *Candida* spp., *Pneumocystis* spp., *Saccharomyces* spp., *Coccidioides* spp., *Aspergillus* spp<sup>1</sup>. This C type lectin, a



Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

transmembrane protein, is the major receptor on macrophages for  $\beta$ 1, 3-glucan, a polymer of glucose present in the fungal cell wall that stimulates phagocytosis and production of inflammatory cytokines<sup>1,3</sup>. Dectin-1 is *N*-glycosylated, a posttranslation modification that contributes to its surface expression and function<sup>4</sup>. The protein contains a single extracellular carbohydrate-recognition domain (CRD) and an immunoreceptor tyrosine activation (ITAM)-like motif in its cytoplasmic tail that is involved in cellular activation<sup>5</sup>.

Dectin-1 is expressed primarily in cells of the innate immune system---monocytes, macrophages, dendritic cells (DC) and neutrophils<sup>1</sup>. The cellular responses observed upon activation of dectin-1 include phagocytosis and an oxidative burst as well as the production of eicosanoids, inflammatory cytokines and chemokines<sup>1,6</sup>. Although dectin-1 has been implicated primarily in the control of fungal infections<sup>2</sup>, other key immunological functions may rely on it as well. Dectin-1 on dendritic cells (DC) appears to engage T lymphocytes through recognition of an unidentified endogenous ligand that binds at a site distinct from the  $\beta$ -glucan binding site<sup>7-8</sup>. Dectin-1 also participates in immunity against mycobacteria and may play a role in mycobacterial phagocytosis<sup>9-11</sup>. Dectin-1 may further contribute to the development of autoimmune diseases as a result of its ability to internalize and present endogenous antigens<sup>12</sup>.

Although the specificity of dectin-1 for the fungal polysaccharide  $\beta$ -glucan has been firmly established by *in vitro* experiments<sup>1,13-16</sup>, most aspects of dectin-1 signaling *in vivo* remain obscure. Dectin-1-mediated responses may require the collaboration with

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

Toll-like receptors (TLR)<sup>17-19</sup>, tetraspanins<sup>20-22</sup> and the DC-SIGN receptor<sup>6</sup>. The association of dectin-1 with potential partners as well as the mechanism by which the initial recognition of glucan is signaled has been hindered by technical difficulties common to membrane proteins that must traffic to the surface of the cell to perform their function. The usual method to label and visualize dectin-1 on the surface of live cells involves the use of fluorophore-conjugated antibodies that bind to epitopes present on dectin-1 or tags added to the protein<sup>6, 21, 23-24</sup>. Such fusions of membrane proteins with GFP or epitope tags can render the protein inactive, unable to transit through the secretory pathway to the surface, or require the use of antibodies that may affect cellular activation or antigen trafficking behavior upon binding to the molecule in question.

Sortagging is a new procedure that finesses these obstacles and allows the assessment of membrane protein partners and their localization. Bacterial sortases can install a variety of small probes on proteins in solution and on the surface of living cells. *Staphylococcus aureus* sortase A recognizes a set of diverse substrates via an LPTXG motif, cleaving the peptide between threonine and glycine and allows efficient and selective labeling with affinity probes or fluorophores<sup>25</sup>. The installation of this small tag provides the opportunity to specifically label a membrane protein such as dectin-1 on the surface of living cells without affecting its cellular location, activity, or interaction with other proteins.

We show here that the application of sortase-mediated transpeptidation indeed allows the installation of a variety of probes on suitably tagged dectin-1 without affecting its

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

location, or activity. This site-specific labeling allowed us to monitor the behavior of functional dectin-1 on the surface of living cells in the presence and absence of a fungal challenge. Using macrophages that express labeled dectin-1 on their cell surface we uncovered an association with galectin-3. Galectin-3 is a PRR that has been reported to recognize carbohydrates uniquely present in the cell wall of *Candida albicans*, triggering a pathogen specific response. The dectin-1/galectin-3 association modulates the induction of TNF- $\alpha$ ; however, a reduction in galectin-3 lowers TNF- $\alpha$  induction when cells are challenged with *C. albicans* but not *S. cerevisiae*. These data suggest that the dectin-1/galectin3 interaction can provide novel insights into the mechanism by which the innate immune system discriminates non-pathogenic from pathogenic fungi. These findings serve as a further illustration of the concept that pattern recognition receptors may modulate their functional properties through association with accessory molecules.

## **Results**

### ***Tagged dectin-1 expressed on mammalian cells is functional and can be labeled through sortagging***

We first determined whether murine dectin-1 equipped with a LPXTG motif near its C-terminus was properly transported to the cell membrane and retained its function. We engineered the sequence encoding the LPETG motif followed by an HA epitope tag at the C-terminus, the extracellular domain of dectin-1. This construct, dectin-1-LPETG-3xHA, was cloned into a retrovirus expression vector and used to obtain stable dectin-1 transductants in HEK 293T and RAW 264.7 cells. Both cell lines expressed readily detectable levels of HA-tagged dectin-1 on the cell surface. After sorting by FACS, we

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

obtained a population where >90% of cells expressed tagged dectin-1 (**Figure A.1a**).

The function of dectin-1 was assessed in stably transduced HEK 293T and RAW 264.7 cells by incubating them with particles of zymosan. Both HEK 293T and RAW 264.7 transfectants but not the empty vector controls showed zymosan-binding, as shown by fluorescence microscopy (**Figure A.1b**) and cytofluorometry (**Figure A.1c**). Although the tags installed on dectin-1 were located at its carbohydrate recognition domain (CRD), the binding of zymosan suggests that the presence of the tag does not interfere with dectin-1's function.

To determine whether sortase A could act on tagged dectin-1 in intact cells, we incubated stably transduced HEK 293T and RAW 264.7 cells in serum-containing medium and exposed them to sortase A together with a biotinylated probe. The intact cells were then lysed, the lysates separated by SDS-PAGE and biotinylated products revealed on blots with streptavidin-horse radish peroxidase (HRP). Indeed, dectin-1 was labeled selectively, because there were no other biotinylated polypeptides that originated either from cell surface or cytosolic proteins. When these lysates were subjected to immunoblotting to detect HA-tagged materials not cleaved by sortase, HA-tagged dectin-1 was present in all samples, both before and after incubation with sortase A and the biotinylated probe (**Figure A.1d**). Even if all dectin-1 were surface exposed, the sortase reaction may not proceed quantitatively, accounting for the presence of HA-positive dectin-1. In addition, any cell-internal dectin-1 would be inaccessible to sortase and so retain its HA tag. Cells that were first lysed, sortagged and only then subjected to

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

streptavidin-HRP immunoblotting also showed the presence of labeled dectin-1. (**Figure A.1d**). We conclude that a fraction of dectin-1 is accessible at the cell surface and amenable to sortase-mediated installation of a biotinylated probe, and that this labeling is highly selective: no biotinylated materials are detected on empty vector controls, and the only biotinylated species detectable in stably transduced cells has the molecular characteristics of dectin-1.

**Sortase-mediated labeling of dectin-1 occurs on the cell surface**

Tagged dectin-1 analyzed by immunoblotting presents itself as a doublet (**Figure A.1d**). Murine dectin-1 is glycosylated<sup>4</sup> and phosphorylated, modifications that could yield multiple electrophoretically distinct species. We sortagged dectin-1 in lysates from stably transduced HEK 293T cells with a biotinylated probe and subjected the samples to digestion with the glycosidases PNGase F or EndoH as well as with Calf Intestinal Alkaline Phosphatase (CIP). PNGase F hydrolyzes all types of *N*-glycan chains, whereas EndoH is specific for high mannose-type *N*-linked glycans. The presence of high mannose-type *N*-linked glycans is usually diagnostic of localization to the ER and/or cis Golgi. After PNGase F treatment, a single biotinylated polypeptide remained; samples treated with either EndoH or CIP still presented as two distinct polypeptides (**Figure A.2**), which must therefore be the result of differential carbohydrate modifications. The exclusive presence of EndoH resistant, biotinylated dectin-1 when intact cells were labeled, indicates that sortase A installed the biotinylated probe only on dectin-1 located on the cell surface.

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

**Sortase-labeled dectin-1 is functional**

To investigate whether sortagged dectin-1 is functional on live cells, we installed an Alexa Fluor647 probe on dectin-1 and determined whether the tagged protein retained its ability to bind zymosan. The specificity of sortagging was evaluated by labeling a mixture of stably transduced dectin-1 HEK 293T cells and control HEK 293T cells transfected with a plasmid encoding GFP. Treatment of the cells with sortase A installed the fluorophore only on the surface of cells that did not express GFP; none of the GFP+ cells was labeled. Furthermore, only those cells that displayed the Alexa647 fluorophore bound zymosan. We conclude that surface-disposed sortagged dectin-1 retains its function (**Figure A.3**).

**Internalization of dectin-1 upon zymosan presentation**

Since sortase A installs a label specifically on surface-disposed dectin-1 but not on cell-internal dectin-1, we could follow the location of sortagged dectin-1 after exposure of stably transduced cells to zymosan. We first installed an Alexa Fluor488 probe on surface-disposed dectin-1 in HEK 293T cells and then exposed the cells to zymosan for 30 minutes. After the 30 minute incubation with zymosan, an Alexa Fluor647 probe was installed on the surface to mark any newly surface exposed dectin-1 that appeared after the first round of sortagging. Cells were then fixed and examined by fluorescence microscopy. The fluorophore installed first showed the presence of dectin-1, either bound to zymosan particles on the surface or internalized. However, dectin-1 sortagged in a second round of labeling with the Alexa Fluor647 was found only on the cell surface (**Figure A.4a**).

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

These data establish that, upon contact with zymosan, dectin-1 is internalized. As expected, regardless of zymosan-induced internalization, dectin-1 continues to emerge at the cell surface via the biosynthetic pathway. To follow dectin-1 internalization in time, a biotinylated probe was installed on dectin-1 in intact HEK 293T cells. The tagged cells were then incubated with cycloheximide to arrest protein synthesis, and the culture was divided in two samples. One sample received zymosan, the other served as a control and both were sampled at different time points. Two different fractions of dectin-1 were assayed: dectin-1 located on the cell surface as inferred from the presence of the biotin label, and dectin-1 inside the cell visualized by means of the HA tag. In the presence of zymosan, the rate of internalization and degradation of biotinylated dectin-1 on the cell surface increases dramatically after 1h of incubation (**Figure A.4b**). The half-life of biotinylated dectin-1 on the cell surface, estimated to be between 8h and 12h in the absence of zymosan, decreases to no more than 2h in the presence of zymosan. The rate of degradation of intracellular HA-tagged dectin-1 was higher than that of biotinylated dectin-1, always showing a rapid decrease after 1h of incubation, regardless of the presence of zymosan (**Figure A.4b**). This observation suggests that a significant fraction of newly synthesized dectin-1 fails to mature and is degraded instead.

**Sortagging identifies an interaction between dectin-1 and galectin-3**

We next explored the possibility of uncovering novel associations between dectin-1 and other potential interactors by using the sortagging technique to install an affinity handle. RAW 264.7 transfectants were incubated with sortase A and a biotinylated probe to label

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

the fraction of tagged dectin-1 located on the cell surface. The cells were lysed and the lysates were then incubated with streptavidin-coated beads. Bound proteins were removed from beads and separated by SDS gel electrophoresis (**Figure A.5**). These proteins were extracted from the gel and subjected to analysis by mass spectrometry. We focused our attention on a protein of about 28 kDa identified as galectin-3 by mass spectrometry. The protein was detected on those samples from transfectants expressing tagged dectin-1. No trace of this protein was observed in any lane corresponding to control cells expressing similarly tagged CD74, an unrelated type II membrane protein.

**Co-immunoprecipitation confirms the interaction between dectin-1 and galectin-3**

We further analyzed the interaction between dectin-1 and galectin-3 in RAW 264.7 macrophages by co-immunoprecipitation. Cell lysates from stably transduced RAW 264.7 cells were incubated with a polyclonal antibody against galectin-3 and analyzed after immunoprecipitation. Western blot analysis using anti-HA antibodies revealed HA-tagged dectin-1, confirming the interaction between galectin-3 and dectin-1 (**Figure A.6a**). We also assayed the effect of zymosan presentation upon this association. Intact stably transduced RAW 264.7 cells were incubated with zymosan particles for 30 min, lysed and then treated with the anti-galectin-3 antibody. As expected, the HA-immunoblotting analysis of the immunoprecipitates showed higher amounts of dectin-1 compared to samples without zymosan presentation. We detected similar results when we lysed stably transduced RAW 264.7 cells and subsequently incubated the lysates with zymosan particles for 30 min, confirming that this response also occurs in the extract. The immunoprecipitates from control cells expressing only HA-tagged CD74 did not



*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

show any HA-tagged protein when analyzed by immunoblotting with anti-HA antibodies (**Figure A.6a**), confirming the specificity of the dectin-1-galectin-3 association.

This specificity of the interaction between dectin-1 and galectin-3 was also confirmed in HEK293T cells stably expressing dectin-1, that had been transfected with a plasmid encoding 3xFLAG-tagged galectin-3. We labeled surface-exposed dectin-1 in intact cells with a biotinylated probe to differentiate the surface disposed dectin-1 from the intracellular HA-tagged fraction. Cells were lysed and a fraction of the lysate was then incubated with streptavidin beads to retrieve biotinylated dectin-1. Bound proteins were removed from beads, separated by SDS gel electrophoresis and analyzed by immunoblotting. When we incubated the blot with an anti-FLAG antibody, we could detect a polypeptide with the expected size for FLAG-tagged galectin-3. We also incubated a part of the cell lysates with an anti-FLAG antibody and analyzed the immunoprecipitates by streptavidin or HA-immunoblotting. Tagged dectin-1 was detected in all cases, confirming the interaction of galectin-3 with the fraction of dectin-1 located on the surface as well as the fraction inside the cell. Neither biotinylated nor HA-tagged CD74 from control cells interacted with galectin-3 (**Figure A.6b**).

**Galectin-3 modulates the induction of TNF- $\alpha$  in association with dectin-1**

To study the biological consequences of the dectin-1-galectin-3 interaction, we performed siRNA-based knockdowns that could lead to changes on the response of RAW 264.7 macrophages after fungal presentation. We analyzed how the reduced expression of galectin-3 could affect the proinflammatory response of macrophages after fungal

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

presentation. The comparison of *S. cerevisiae* with *C. albicans* requires that both be in the yeast form. However *C. albicans* produces filaments under the conditions of this assay. For this reason we inactivated the fungi with UV (both *C. albicans* and *S. cerevisiae*) because UV-irradiated *Candida* cells do not present as filaments and remain in the yeast form. We first incubated galectin-3-competent RAW 264.7 cells with UV-treated *S. cerevisiae* or *C. albicans* at a ratio of 1:5 and measured TNF- $\alpha$  mRNA expression after 4h. There was no significant increase in the level of induction of TNF- $\alpha$  expression, compared to cells without fungal presentation. These cells have very low expression levels of dectin-1 (**Figure A.7a**), so we next studied the response in cells that stably express dectin-1. In these cells we observed a significant increase in TNF- $\alpha$  mRNA levels after fungal presentation ( $p < 0.05$ ). This observation confirms the key role of dectin-1 in the induction of TNF- $\alpha$  when confronted with either *S. cerevisiae* or *C. albicans*. We then studied the response in RAW 264.7 cells where the level of galectin-3 was reduced by an RNA hairpin directed against galectin-3 mRNA. This hairpin successfully reduced galectin-3 mRNA (**Figure A.7a**). Cells with and without the hairpin had similar levels of TNF- $\alpha$  when they were incubated with *S. cerevisiae*. However, when cells were incubated with *C. albicans* there is a nearly fourfold decrease ( $p < 0.05$ ) compared to galectin-3-competent cells (**Figure A.7b**). These data show that galectin-3 in concert with dectin-1 differentially modulates the response to these two fungi. Significantly, the intensity of the response is much greater for the pathogen than the non-pathogen.

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

**Discussion**

In this report we show that dectin-1 suitably modified for sortagging is active in live macrophages; it transits through the secretory pathway and is inserted into the plasma membrane, where it recognizes its normal ligand,  $\beta$ -glucan and initiates the signal for an inflammatory response when challenged with fungi. Cells expressing tagged dectin-1 were able to recognize and bind zymosan, a particulate extract from *S. cerevisiae* cell walls that consist of protein-carbohydrate complexes with high amounts of  $\beta$ -glucan. This binding was observed in stably transduced HEK 293T cells, a cell line that does not show phagocytic activity unless dectin-1 is expressed<sup>26</sup>. This ability of tagged dectin-1 to confer binding and internalization of zymosan in otherwise non-phagocytic cells is similar to that reported NIH-3T3 fibroblasts transfected with dectin-1<sup>14, 23, 27</sup>, and suggests that the tagged protein retains this function as well.

Sortagging revealed that a fraction of dectin-1 exposed at the cell surface is accessible to sortase-mediated labeling in a very specific fashion. This finding agrees with previous studies where the efficiency and specificity of the method were tested in complex protein mixtures, including mixtures of soluble proteins, and on live cells<sup>25</sup>. Immunoblotting of surface-exposed dectin-1 identified two distinct polypeptides, the consequence of differential glycosylation. Dectin-1 requires *N*-linked glycosylation, since its suppression affects cell surface expression of dectin-1 and is essential for the recognition of fungal  $\beta$ -glucan<sup>4</sup>.

Site-specific installation of fluorophores on dectin-1 by means of sortagging permitted

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

visualization of the receptor on live cells without compromising its ability to recognize  $\beta$ -glucan. This fluorophore-labeled dectin-1 binds zymosan on the cell surface and is subsequently localized intracellularly. This internalization could be monitored in time by installing a biotinylated probe on surface-exposed dectin-1. Upon recognition of  $\beta$ -glucan, the levels of biotinylated dectin-1 from the cell surface decrease dramatically after 1h of incubation. By contrast in the absence of zymosan, dectin-1 was more stable on the cell surface. The fraction of dectin-1 inside the cell does not behave differently with or without zymosan and appeared to be degraded at comparable rates. The degradation rate of intracellular dectin-1 was always higher than the surface-disposed fraction. This observation suggests that a fraction of newly synthesized, cell-internal dectin-1 fails to mature and is degraded. This retention may be similar to that reported for other glycoproteins such as the KIR3DL1 NK cell receptor, where only a minor fraction escapes the ER<sup>28</sup>, while the surface-exposed fraction is fully functional. The internalization of dectin-1 upon its interaction with ligand may attenuate pathways involved in induction of the innate immune response<sup>29</sup>. Once dectin-1 is internalized, the route of intracellular processing may well depend on the nature of the ligand. For example, dectin-1 is directed to lysosomes during uptake of zymosan, but recycled to the membrane during uptake of the soluble ligand laminarin<sup>23</sup>.

Installation of a biotinylated probe on surface-exposed dectin-1 and subsequent pull-down with streptavidin beads revealed a previously undiscovered association between dectin-1 and galectin-3 in RAW 264.7 macrophages. Mass spectrometry of proteins recovered by affinity adsorption and separated by electrophoresis identified galectin-3 as

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

a potential interacting protein. The association between dectin-1 and galectin-3 was confirmed by co-immunoprecipitation in RAW 264.7 cells and also in HEK 293T transfectants expressing both proteins. This association occurred on the cell surface as well as inside the cell.

The association between dectin-1 and galectin-3 connects two of the key signature molecules on the surface of fungi,  $\beta$ -glucan and oligomannans. Dectin-1 recognizes  $\beta$ -glucan, a carbohydrate that accounts for a major fraction of the cell wall of most fungi and it is covalently linked to an outer layer of mannoproteins<sup>1, 15-16</sup>. Galectin-3 has been described as a PRR that recognizes specific  $\beta$ -1, 2 oligomannans from the cell wall of certain fungi<sup>30-31</sup>. Previous work confirmed that galectin-3 was able to bind to four *Candida* species expressing different combinations of  $\beta$ -1, 2-oligomannans, but did not bind *S. cerevisiae*, which lacks this specific oligosaccharide in its cell wall. Moreover galectin-3 binding to *C. albicans* was directly fungicidal in a complement-independent fashion<sup>31</sup>. There are specialized PRRs such as TLRs, DC-SIGN or the mannose receptor (MR) that also recognize mannose exposed at the cell wall<sup>32</sup>; however, only galectin-3 has been reported to discriminate between pathogenic and non-pathogenic fungi that lack specific  $\beta$ -linked mannosides<sup>33</sup>.

Consistent with the role of association between these molecules in fungal recognition, we found that galectin-3 modulates the inflammatory response of macrophages in cooperation with dectin-1. In accordance with previous reports<sup>17</sup>, the increase in the levels of dectin-1 expression in RAW 264.7 macrophages presented with fungi leads to

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

an increase of TNF- $\alpha$  expression. However, reduction of galectin-3 expression in these cells results in a differential change in the inflammatory response depending upon which fungal cell is presented to the macrophages. Cells with reduced galectin-3 still induce TNF- $\alpha$  upon a challenge with *S. cerevisiae* but they fail to generate the same TNF- $\alpha$  induction after incubation with *C. albicans*. This result suggests that the dectin-1/galectin3 association plays an important role in the recognition of *C. albicans* that is distinct from that of *S. cerevisiae*.

Our finding of differential recognition of *S. cerevisiae* and *C. albicans* is supported by previous work on galectin-3 knock-out mice. Galectin-3 associates with TLR2 at the cell membrane. Jouault et.al.<sup>33</sup> reported co-precipitation of galectin-3 and TLR2 in THP-1 cells treated with PMA. TLR2 recognizes phospholipomannan and activates inflammatory genes that trigger cytokine production<sup>34-35</sup>. The association with galectin-3 appears to modulate this TLR2-driven inflammatory response. Macrophages from galectin-3 knock-out mice resulted in lower levels of TNF- $\alpha$  in response to *C. albicans* as compared with cells from wild type mice whereas there was no significant difference in TNF- $\alpha$  induction between wild type and mutant when cells were incubated with *S. cerevisiae*<sup>33</sup>. Galectin-3 seems to display a similar role in association with TLR2, so, it is likely that the association of galectin-3 with different cell surface receptors is a mechanism that would allow the cell to adjust the appropriate inflammatory response toward a specific fungal challenge, primarily after sensing *C. albicans*. These findings taken together with our results suggest that the association between galectin-3 and dectin-1 is key to modulating the proinflammatory response that distinguishes between

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

pathogenic and non-pathogenic fungi.

## **Methods**

### **Reagents**

Zymosan and protease inhibitor cocktail were from Sigma-Aldrich. The anti-HA (hemagglutinin) antibody conjugated with R-Phycoerythrin (R-PE) and the IgG isotype control antibody were from Columbia Biosciences. The anti-HA antibody conjugated with peroxidase was from Roche Applied Science. The streptavidin conjugated with horse radish peroxidase (HRP) was from GE Healthcare. The polyclonal antibody against galectin-3 was from Santa Cruz Biotechnology. Anti-FLAG antibody conjugated with HRP and anti-FLAG-coated magnetic beads were from Sigma-Aldrich. The monoclonal antibody to detect  $\beta$ -actin was from Abcam. Immunoprecipitation Kit-Dynabeads Protein G and Dynabeads Streptavidin C1 were purchased from Invitrogen. EndoH, PNGase F and Calf Intestinal Alkaline Phosphatase were from New England BioLabs.

Cyclohexamide was from EMD Biosciences. Alexa Fluor 488/555/647 carboxylic acid, succinimidyl esters were from Invitrogen Molecular Probes. Transient transfections were done with FuGENE 6 Transfection Reagent from Roche Applied Science. For all experiments quantitation of protein content was performed using BCA protein assay kit from Thermo Fisher Scientific.

### **Synthesis of GGGK-A647, GGGK-A488 and biotinylated probes**

The GGGK peptide was manually synthesized by standard Fmoc-based solid phase peptide synthesis protocols on Rink Amide Resin (Novabiochem). The Fmoc-protected

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

peptide was cleaved from the resin by treatment with 2.5 ml of 95:3:2 TFA-TIPS/H<sub>2</sub>O (5X, 15 min each), which also removed the 4-methyltrityl (Mtt) protecting group on the lysine residue. The combined cleavage solutions were concentrated, dissolved in methanol, and precipitated with cold diethyl ether.

Coupling of the Fmoc-protected peptide to the amine reactive Alexa Fluor dyes was accomplished in solution. One equivalent of Fmoc-GGGK peptide was mixed with 0.5 equivalents of either Alexa Fluor 647 carboxylic acid, succinimidyl ester or Alexa Fluor 488 carboxylic acid, succinimidyl ester, and 4 equivalents of DIPEA (Sigma-Aldrich) in anhydrous DMSO and incubated at room temperature for 6 h. Under these conditions, Fmoc deprotection also occurred.

The Alexa488 peptide was purified by reversed-phase HPLC on a Waters Delta Pak 15 mm, 100 Å C18 column (7.8 x 300 mm, MeCN:H<sub>2</sub>O gradient mobile phase containing 0.1% TFA, 3 ml/min) and the Alexa647 peptide was purified on a Waters 5PE column (8 x 250 mm, MeCN:H<sub>2</sub>O gradient mobile phase with 0.1% TFA, 3ml/min). Two peaks were collected for the Alexa488 peptide, presumably representing different isomers. Fractions were lyophilized and dissolved in water.

Peptide identity was confirmed for the Alexa488 peptide by MALDI-TOF MS (matrix sinapinic acid), [M<sup>+</sup>] = 831.16, obs = 832.862; both HPLC peaks contained this mass. We were unable to observe the Alexa647 peptide by MALDI-TOF, so we inferred the molecular weight and activity as a nucleophile by setting up a test transpeptidation



*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

reaction with sortase A, and an LPETG tagged GFP substrate as described (Antos et al., 2008), and observing the mass change in the transpeptidation product by ESI-MS on a Micromass LCT mass spectrometer (Micromass® MS Technologies) and a Paradigm MG4 HPLC system equipped with a HTC PAL autosampler (Michrom BioResources) and a Waters Symmetry 5 mm C8 column (2.1x 50 mm, MeCN:H<sub>2</sub>O (0.1% formic acid) gradient mobile phase, 150 ml/min). The predicted molecular weight for the Alexa647 peptide is 1155.06, obs=1155.0.

The biotinylated probe was synthesized as previously described (Popp et al., 2007), except the peptide scaffold consisted of GGGK.

**Stable Cell lines and siRNA**

The coding region for murine dectin-1 was cloned into the retroviral vector pLHCX (Clontech Laboratories) and tagged at the C-terminus with additional coding region for the following residues: GGGGSGGGGSLPETG. This tag includes a flexible linker and the recognition site for sortase A. A 3xHA epitope tag was also added C-terminal to the sortase recognition site. Retrovirus was prepared by transfecting HEK 293T cells with plasmids encoding VSV-G and Gag-Pol as well as the dectin-1 pLHCX construct.

Medium was collected 24 and 48 hours post-infection, filtered through a 0.45 µm membrane, and added to RAW264.7 and HEK 293T cells (American Type Culture Collection, ATCC TIB-71 and ATCC CRL-1573) with 8 µg/ml polybrene. Cells were spun at 1000xg for 90 min. Media was changed and cells were allowed to recover for 24 h before re-infection and selection to obtain Hygromycin-resistant populations. Cells

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

were sorted on a FACS ARIA cell sorter (BD Biosciences) to typically >90% purity on post-sort analysis. A subset of RAW 264.7 and HEK 293T cells were transduced with empty retroviral plasmid. The Hygromycin-resistant population obtained was used as a control for further experiments. Cells were cultured under Hygromycin B selection (250 µg/ml in RAW 264.7 cells and 125 µg/ml in HEK 293T cells) in complete DMEM/10% FBS supplemented with penicillin 100 units/ml, and streptomycin 100 mg/ml.

To elaborate galectin-3 knockdowns, the retroviral vector pKLO.1 was used for that purpose. We used three different short hairpin RNAs (shRNA) with the sequences: 5'-AGAGTCATTGTGTGTAACACG-3'; 5'-AACCATCGGATGAAGAACCTC-3' and 5'-AGCTGCCTGTCTTTATATGCC-3'. RAW 264.7 cells with stable dectin-1 expression were infected with the shRNAs as previously described and puromycin-resistant clones were selected. RAW 264.7 cells and dectin-1-stably-transfected RAW 264.7 cells were transduced with shRNA directed against GFP mRNA (5'GCCACAACATCGAGGACGGCA-3') and used as a control for further experiments. Cells were cultured under puromycin selection (8 µg/ml).

### **Cell Sortagging**

Cells were cultured in complete DMEM/10% FBS and incubated in a 6-well plate in a 5% CO<sub>2</sub> humidified incubator at 37°C. The cells were harvested by scraping in ice-cold PBS. After centrifugation, the cell pellet was resuspended in complete DMEM media. Tagged dectin-1 was labeled by incubating the cells with sortase A (200 µM) and the probe to be installed: Biotinylated probe (500 µM), Alexa488 probe (10 µM) or

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

Alexa647 probe (10  $\mu$ M). The cells were incubated at 37°C for 30 min. As a control, a subset of cells was incubated with each probe but without sortase enzyme.

For the cell lysate sortagging, cells were lysed with the addition of PBS (Phosphate Buffered Saline) supplemented with 1% Nonidet P-40 (NP-40) and protease inhibitor cocktail to the pellet. The lysate (10  $\mu$ g) was incubated with sortase A (200  $\mu$ M) and the biotinylated probe (500  $\mu$ M) in sortase buffer (50mM Tris pH 7.5, 150mM NaCl, 10mM CaCl<sub>2</sub>) at 37°C for 30 min. The whole reaction was analyzed by western blot. A fraction of the lysate was incubated with no enzyme and used as a control.

**Glycosidase and phosphatase treatment**

Cultured cells were sortagged with a biotinylated probe and centrifuged. The cells pellet was lysed with PBS supplemented with 1% NP-40 and protease inhibitor cocktail. The removal of carbohydrate residues was performed incubating samples from cell lysates (10  $\mu$ g) with 1,500 units of Endoglycosidase H (EndoH) or N-Glycosidase F (PNGase F) as described by the manufacturer. The cell lysates were incubated in glycoprotein denaturing buffer at 100°C for 10 min. Then, 10x G5 Reaction Buffer was added for the EndoH digestion, and 10x G7 Reaction Buffer and 10% NP-40 were added for the PNGase F treatment. The alkaline phosphatase digestion was carried out as previously described (Qi et al, 2006). Thirty units of Calf Intestinal Alkaline Phosphatase (CIP) were used to digest 10  $\mu$ g from cell lysates at 37°C for 1h. HEK 293T cells transfected with CD74 were used as a positive control. The CD74 protein contained two mutated leucines (L7A, L17A) that prevent recycling of the invariant chain and allows stable, detectable

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

levels of the protein on the cell surface.

**Protein degradation assay**

Protein degradation assay was performed as described previously (Fu et al., 2007). HEK 283T cells were cultured on 12-well plates and sortagged with a biotinylated probe.

Cyclohexamide (50µg/ml) was added to medium right after sortagging, and the cells were harvested at different time points (0, 15min, 30min, 1h, 2h, 4h, 8h, and 12h). A fraction of the cell lysate was digested with EndoH for 2h. All cell lysates in SDS sample buffer (raw lysate and EndoH-treated fraction) were boiled and then analyzed by western blotting using streptavidin, or anti-HA antibody or anti-β-actin antibody. All samples were incubated under two different conditions, with and without zymosan. Ten particles of zymosan/cell were added to the culture and incubated for 30 min at 37°C. The quantification of the levels of protein from the immunoblots was performed with ImageJ software (NIH).

**Streptavidin beads affinity precipitation and co-immunoprecipitation**

RAW 264.7 macrophages expressing either tagged dectin-1 or tagged CD74 (with two mutated leucines-L7A, L17A) were sortagged with a biotinylated probe. Transfectants were then divided in two fractions: one sample received zymosan, (10 particles/cell) for 30 min at 37°C, the other one did not. Cells were lysed with 0.5% NP-40 and a protease inhibitor mix, and precleared with magnetic beads (without streptavidin) for 1h at 4°C and then incubated with magnetic streptavidin Dynabeads overnight at 4°C. Beads were washed three times in PBS with 0.5% NP-40, and bound protein was eluted with SDS-

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

PAGE sample buffer. The samples were then separated by SDS-PAGE. The gels were immunoblotted with streptavidin or the appropriate antibody or stained with silver stain, and bands cut for analysis by mass spectrometry.

For the co-immunoprecipitation assay RAW 264.7 or HEK 293T transfectants were lysed and the lysates precleared with IgG and Dynabeads/protein G for 1 h at 4°C. Anti-Galectin-3 antibody (5µg) was immobilized on 50mL Dynabeads/protein G and incubated with 500µg of cell extract supernatant overnight at 4°C. The immunoprecipitates were washed three times in PBS with 0.5% NP-40 and eluted in 30µl of gel sample buffer containing SDS. For anti-FLAG immunoprecipitation, HEK 293T cells stably expressing tagged dectin-1 were cultured overnight on a 6-well plate and transfected with a vector encoding 3xFLAG-tagged galectin-3. Cell lysates were precleared with Dynabeads (without streptavidin or antibody attached) and incubated overnight at 4°C with anti-FLAG-coated magnetic beads. For immunoblotting, the appropriate antibodies or streptavidin were used.

**Ligand binding assays**

HEK 293T cells were plated onto a poly-lysine coated chambered coverglass at a density of  $3 \times 10^5$  cells/well ( $0.7 \text{ cm}^2$ ), and allowed to attach overnight. Zymosan was labeled with Alexa Fluor488, Alexa Fluor555 or Alexa Fluor647 carboxylic acid, succinimidyl ester following manufacturer instructions. Cells were incubated with labeled zymosan (10 particles/cell) 30 min at 37°C and then washed thoroughly with PBS to remove unbound

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

particles. For binding assays on sortagged cells, cells were sortagged on the chambered coverglass in DMEM/10% FBS media, and then washed thoroughly with PBS. Labeled cells were incubated with zymosan in DMEM/10% FBS media 30 min at 37°C and then washed with PBS. If a second fluorophore was installed through sortagging, at that stage cells were labeled again on the coverglass with a different fluorophore and then washed with PBS. For the binding assay on a mixed population of HEK 293T cells, cells were cultured overnight on a 6-well plate and transfected with 2µg of pCAG-GFP vector (Matsuda and Cepko, 2004). Cells were incubated with the transfection reagent overnight at 37°C. After the incubation, 3 x 10<sup>3</sup> GFP-transfected cells and 3 x 10<sup>3</sup> stably dectin-transfected cells were transferred to each of the wells of a poly-lysine coated chambered coverglass. The ligand binding assay was performed as described.

**RNA extraction and RT-qPCR**

Total RNA was extracted from cells using Rneasy (Qiagen). To measure the relative mRNA levels of the different genes, quantitative RT-PCR (qPCR) was done using SYBR Green (Applied Biosystems) and a 7500 real-time PCR system (Applied Biosystems), following the manufacturing protocol. Each sample was tested in triplicate. GAPDH was used as the invariant control. The following primer sets were used to detect dectin-1, galectin-3 and mTNF-α. Dectin-1: 5'-ACCACAAGCCCACAGAATCATC-3' (forward), 5'-CATGGTCCAATTAGGAAGGCAA-3' (reverse); galectin-3: 5'-TATCCTGCTGCTGGCCCTTAT-3' (forward), 5'-CACTGTGCCCATGATTGTGATC-3' (reverse); mTNF-α: 5'-ATGGCCTCCCTCTCATCAGTTC-3' (forward), 5'-TTGGTGGTTTGCTACGACGTG-3' (reverse). For determination of TNF-alpha mRNA

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

levels, cells were incubated with UV-inactivated *S. cerevisiae* or *C. albicans* at a 1:5 ratio or without fungi for 4 h at 37°C. The UV inactivation of yeast was performed as previously described (Wheeler et al., 2006). The equivalent of  $2.5 \times 10^7$  cells resuspended in 1ml of PBS were exposed to four doses of 100,000  $\mu\text{joules}/\text{cm}^2$  in a CL-100 UV-crosslinker (UVP, LLC). The experiment was repeated three times in different days.

### **Western blotting**

When live sortagged cells were analyzed, cells were centrifuged and the pellet was lysed with PBS supplemented with 1% SDS or 0.5-1% NP-40 and protease inhibitor cocktail. For western blotting, a fixed amount of total protein was mixed with sample buffer (50mM Tris, pH 7.5, 2% SDS, 10% glycerol, 2.5% 2-mercaptoethanol, 0.02% Bromophenol Blue), and resolved by Tris-glycine SDS-PAGE. After transferring to PVDF membrane, proteins were analyzed by immunoblotting. HRP-conjugated antibodies or streptavidin were used at a 1:10,000 dilution.

### **Fluorescence Microscopy**

Cells were sortagged and/or incubated with Alexa Fluor647-labeled or Alexa Fluor555-labeled zymosan and fixed with 3% paraformaldehyde. Samples were visualized on chambered coverglass with an inverted Nikon TE2000-s microscope (Nikon) equipped with Spot RT Camera (Diagnostic Instruments). The imaging camera was set to capture 8-bit images that were subsequently processed with Photoshop (Adobe Systems). Samples were examined with a 40x/0.75 M/N2 dry objective lens at room temperature.

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

**Flow cytometry (FACS)**

Cells were fixed with 3% paraformaldehyde. Fluorescence was quantified on a FACS caliber cytometer (BD Biosciences). Cells were gated by forward and side scatter based on wild-type cell size and shape, and mean fluorescence intensity of 10,000 labeled cells was calculated using Cellquest software (BD Biosciences).

**Statistical analysis**

Data from RT-qPCR were acquired from three independent experiments performed on multiple days. Statistical significance was determined with two-tailed Student's T-test. Values of  $p < 0.05$  were considered significant.

**Acknowledgements**

The authors thank Dr. Valmik Vyas for the precious gift of galectin-3 hairpin RNAs, Dr Matsuda for the pCAG-GFP plasmid and Eric Spooner for his technical assistance with mass spectrometry. This work was funded by the National Institutes of Health (NIH) grant GM040266. Additional funding was provided from the Fulbright/Spanish Ministry of Education and Science Visiting Scholar Program (FU2006-0983).



Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

**References**

1. Brown, G.D. Dectin-1: a signalling non-TLR pattern-recognition receptor. *Nat Rev Immunol* **6**, 33-43 (2006).
2. Sun, L. & Zhao, Y. The biological role of dectin-1 in immune response. *Int Rev Immunol* **26**, 349-364 (2007).
3. Brown, G.D. & Gordon, S. Immune recognition. A new receptor for beta-glucans. *Nature* **413**, 36-37 (2001).
4. Kato, Y., Adachi, Y. & Ohno, N. Contribution of N-linked oligosaccharides to the expression and functions of beta-glucan receptor, Dectin-1. *Biol Pharm Bull* **29**, 1580-1586 (2006).
5. Fuller, G.L. et al. The C-type lectin receptors CLEC-2 and Dectin-1, but not DC-SIGN, signal via a novel YXXL-dependent signaling cascade. *J Biol Chem* **282**, 12397-12409 (2007).
6. Valera, I. et al. Costimulation of dectin-1 and DC-SIGN triggers the arachidonic acid cascade in human monocyte-derived dendritic cells. *J Immunol* **180**, 5727-5736 (2008).
7. Willment, J.A., Gordon, S. & Brown, G.D. Characterization of the human beta-glucan receptor and its alternatively spliced isoforms. *J Biol Chem* **276**, 43818-43823 (2001).
8. Grunebach, F., Weck, M.M., Reichert, J. & Brossart, P. Molecular and functional characterization of human Dectin-1. *Exp Hematol* **30**, 1309-1315 (2002).
9. Yadav, M. & Schorey, J.S. The beta-glucan receptor dectin-1 functions together with TLR2 to mediate macrophage activation by mycobacteria. *Blood* **108**, 3168-3175 (2006).
10. Rothfuchs, A.G. et al. Dectin-1 interaction with Mycobacterium tuberculosis leads to enhanced IL-12p40 production by splenic dendritic cells. *J Immunol* **179**, 3463-3471 (2007).
11. Shin, D.M. et al. Mycobacterium abscessus activates the macrophage innate immune response via a physical and functional interaction between TLR2 and dectin-1. *Cell Microbiol* **10**, 1608-1621 (2008).
12. Reid, D.M., Gow, N.A. & Brown, G.D. Pattern recognition: recent insights from Dectin-1. *Curr Opin Immunol* **21**, 30-37 (2009).
13. Brown, G.D. et al. Dectin-1 is a major beta-glucan receptor on macrophages. *J Exp Med* **196**, 407-412 (2002).
14. Herre, J., Gordon, S. & Brown, G.D. Dectin-1 and its role in the recognition of beta-glucans by macrophages. *Mol Immunol* **40**, 869-876 (2004).
15. Palma, A.S. et al. Ligands for the beta-glucan receptor, Dectin-1, assigned using "designer" microarrays of oligosaccharide probes (neoglycolipids) generated from glucan polysaccharides. *J Biol Chem* **281**, 5771-5779 (2006).
16. Adams, E.L. et al. Differential high-affinity interaction of dectin-1 with natural or synthetic glucans is dependent upon primary structure and is influenced by polymer chain length and side-chain branching. *J Pharmacol Exp Ther* **325**, 115-123 (2008).

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

17. Brown, G.D. et al. Dectin-1 mediates the biological effects of beta-glucans. *J Exp Med* **197**, 1119-1124 (2003).
18. Underhill, D.M. Collaboration between the innate immune receptors dectin-1, TLRs, and Nods. *Immunol Rev* **219**, 75-87 (2007).
19. Ferwerda, G., Meyer-Wentrup, F., Kullberg, B.J., Netea, M.G. & Adema, G.J. Dectin-1 synergizes with TLR2 and TLR4 for cytokine production in human primary monocytes and macrophages. *Cell Microbiol* **10**, 2058-2066 (2008).
20. Mantegazza, A.R. et al. CD63 tetraspanin slows down cell migration and translocates to the endosomal-lysosomal-MIICs route after extracellular stimuli in human immature dendritic cells. *Blood* **104**, 1183-1190 (2004).
21. Meyer-Wentrup, F. et al. Dectin-1 interaction with tetraspanin CD37 inhibits IL-6 production. *J Immunol* **178**, 154-162 (2007).
22. van Spruiel, A.B. et al. The tetraspanin protein CD37 regulates IgA responses and anti-fungal immunity. *PLoS Pathog* **5**, e1000338 (2009).
23. Herre, J. et al. Dectin-1 uses novel mechanisms for yeast phagocytosis in macrophages. *Blood* **104**, 4038-4045 (2004).
24. Kennedy, A.D. et al. Dectin-1 promotes fungicidal activity of human neutrophils. *Eur J Immunol* **37**, 467-478 (2007).
25. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**, 707-708 (2007).
26. Gantner, B.N., Simmons, R.M. & Underhill, D.M. Dectin-1 mediates macrophage recognition of *Candida albicans* yeast but not filaments. *EMBO J* **24**, 1277-1286 (2005).
27. Heinsbroek, S.E. et al. Expression of functionally different dectin-1 isoforms by murine macrophages. *J Immunol* **176**, 5513-5518 (2006).
28. Taner, S.B. et al. Interactions of NK cell receptor KIR3DL1\*004 with chaperones and conformation-specific antibody reveal a functional folded state as well as predominant intracellular retention. *J Immunol* **186**, 62-72 (2011).
29. Hernanz-Falcon, P., Joffre, O., Williams, D.L. & Reis e Sousa, C. Internalization of Dectin-1 terminates induction of inflammatory responses. *Eur J Immunol* **39**, 507-513 (2009).
30. Fradin, C., Poulain, D. & Jouault, T. beta-1,2-linked oligomannosides from *Candida albicans* bind to a 32-kilodalton macrophage membrane protein homologous to the mammalian lectin galectin-3. *Infect Immun* **68**, 4391-4398 (2000).
31. Kohatsu, L., Hsu, D.K., Jegalian, A.G., Liu, F.T. & Baum, L.G. Galectin-3 induces death of *Candida* species expressing specific beta-1,2-linked mannans. *J Immunol* **177**, 4718-4726 (2006).
32. Willment, J.A. & Brown, G.D. C-type lectin receptors in antifungal immunity. *Trends Microbiol* **16**, 27-32 (2008).
33. Jouault, T. et al. Specific recognition of *Candida albicans* by macrophages requires galectin-3 to discriminate *Saccharomyces cerevisiae* and needs association with TLR2 for signaling. *J Immunol* **177**, 4679-4687 (2006).

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

34. Jouault, T. et al. Candida albicans phospholipomannan is sensed through toll-like receptors. *J Infect Dis* **188**, 165-172 (2003).
35. Underhill, D.M. et al. The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811-815 (1999).

## Figure Legends

**Figure A.1. Tagged dectin-1 is expressed on the cell surface of mammalian cells, is functional and can be labeled through sortagging.**

(a) HEK 293T and RAW 264.7 cells were transduced with a retroviral expression vector containing murine tagged dectin-1 (with LPETG motif and 3xHA tag at the C-terminus). Cells were then sorted by FACS to obtain homogeneous populations of cells expressing readily detectable levels of the tagged protein. Tagged dectin-1 was detected with anti-HA antibodies conjugated with R-PE (anti-HA). A sample of cells was incubated with IgG isotype control antibody (isotype).

(b,c) Tagged dectin-1 containing the LPETG motif at the C-terminus is still functional. Stably transduced HEK 293T and RAW 264.7 cells (Dectin +) were incubated on chambered coverglass with Alexa Fluor647 (red) or Alexa Fluor488 (green)-labeled particles of zymosan (10 particles/cell) for 30 min. Tagged dectin-1 recognized  $\beta$ -glucan and bound zymosan particles as shown by fluorescence microscopy (B) and cytofluorometry (C). Transfectants with an empty retrovirus plasmid (Dectin -) were used as a control.

(d) Sortase A can recognize tagged dectin-1 in intact mammalian cells and install a biotinylated probe. HEK 293T and RAW 264.7 transductants were incubated with sortase A together with a biotinylated probe. Cell lysates from the sortagged cells were analyzed by immunoblotting on SDS-PAGE. The only detectable biotinylated species has the molecular characteristics of dectin-1 (Biotin). HA immunoblotting showed persistence of HA-tagged dectin-1 (HA tag) in all samples both before and after incubation with sortase

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

A and the biotinylated probe. 1: intact cells were lysed after sortagging; 2: cells were first lysed and lysates were subsequently sortagged.

**Figure A.2. Sortagging of dectin-1 occurs on the surface of intact cells.**

Stably transduced HEK 293T cells (293dectin) were sortagged with a biotinylated probe, lysed and treated with PNGase F, EndoH or CIP. Streptavidin-HRP immunoblotting reveals deglycosylated dectin-1 when treated with PNGase F, showing that the biotinylated probe was installed specifically on tagged dectin-1 on the cell surface (Biotin). Anti-HA blotting shows deglycosylation of dectin-1 when treated with either PNGase F or EndoH (HA tag). HEK 293T cells transfected with tagged CD74 were used as a positive control (293CD74).

**Figure A.3. A fluorophore-tagged dectin-1 retains function.**

A mixed population of stably transduced HEK 293T cells expressing tagged dectin-1 and control HEK 293T cells transfected with GFP were incubated with sortase A and the Alexa Fluor647 (red) probe. Afterwards, the population was incubated with Alexa Fluor555 (orange)-labeled particles of zymosan (10 particles/cell) for 30 min. Red-labeled sortagged dectin-1 recognized  $\beta$ -glucan and bound Alexa Fluor555-labeled zymosan particles (arrowheads). Control GFP<sup>+</sup> cells without sortase treatment and incubated with the fluorophore showed no red fluorescence.

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

**Figure A.4. Dectin-1 can be visualized by sortagging either on the cell surface or internalized upon zymosan binding and zymosan increases the degradation rate of dectin-1 located on the cell surface.**

(a) HEK 293T cells were incubated with sortase A and the Alexa Fluor488 (green) probe, then incubated with Alexa Fluor555-labeled (orange) zymosan (10 particles/cell) for 30 min. After zymosan presentation, cells were incubated with sortase A and the Alexa Fluor647 (red) probe. Fluorophore-labeled dectin-1 appears bound to Alexa Fluor555-labeled (orange) zymosan particles (arrowheads, Alexa488/555 overlay). Alexa Fluor647-labeled (red) dectin-1 appears only on the cell surface. Alexa Fluor488-labeled (green) dectin-1 both appears on the cell surface or internalized (arrowheads) (Alexa488/647 overlay).

(b) A biotinylated probe was installed on stably transfected HEK 293T cells through sortagging. The cells were then incubated in the presence of cyclohexamide (50 $\mu$ g/mL) to inhibit protein synthesis. Cells were divided in two groups, one with and the other without zymosan. Samples were collected at different time points and levels of biotinylated dectin-1 on the cell surface and HA-tagged dectin-1 inside the cell were analyzed by immunoblotting. In the absence of zymosan, the levels of biotinylated dectin-1 from the cell surface decrease gradually after 12h of incubation to near 50% of the initial levels (ZYM-). However, upon addition of zymosan, the levels of biotinylated dectin-1 decrease rapidly after 1h of incubation due to internalization and degradation of the dectin-1/zymosan complex (ZYM+). The degradation rate of the intracellular HA-tagged dectin-1 in cytoplasm, Golgi or ER (HA tag, EndoH) did not change regardless of the presence of zymosan. Biotin: biotinylated dectin-1; HA: HA-tagged dectin-1; EndoH:

Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

EndoH-treated HA-tagged dectin-1; dect-: control cells with no tagged-dectin-1 expression.

**Figure A.5. Sortagging reveals an interaction between dectin-1 and galectin-3.**

Dectin-1 from RAW 264.7 stably transduced cells was labeled with a biotinylated probe. A fraction of these cells was incubated with zymosan particles (10 particles/cell) for 30 min at 37°C. Cell lysates were precleared and incubated with magnetic streptavidin Dynabeads overnight at 4°C. Beads were washed three times in PBS with 0.5% NP-40, and bound protein was eluted with SDS-PAGE sample buffer. The samples were then separated by SDS-PAGE.

(a) The gel was incubated with silver stain, and the bands extracted for analysis by mass spectrometry (MS). The site where dectin-1 (d), galectin-3 (g) and the control CD74 (c) were localized in the gel after MS analysis is indicated.

(b) Western analysis with HRP-conjugated streptavidin reveals biotinylated dectin-1 on sortagged cells before and after precipitation with streptavidin beads (Biotin). RAW 264.7 cells expressing tagged CD74 were used as a control. dect-: cells with no expression of tagged dectin-1 or tagged CD74; dt: cells expressing tagged dectin-1; ct: control cells expressing tagged CD74; zym: zymosan.

**Figure A.6. Co-immunoprecipitation of dectin-1 with galectin-3.**

(a) RAW 264.7 stably transduced cells were split in three different groups. One group was lysed with 0.5% NP-40 (1). A second group was incubated with zymosan (10 particles/cell) for 30min at 37°C and then lysed (2). A third group was lysed and then

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

incubated with zymosan for 30 min at 37°C (3). Cell extracts (500µg) were precleared and incubated with anti-galectin-3 antibody (5 µg) overnight at 4°C. Beads were washed three times in PBS with 0.5% NP-40, and boiled with SDS-PAGE sample buffer to elute bound proteins. The samples were then separated by SDS-PAGE and analyzed by immunoblotting. On the left are immunoblots of just the lysate showing that under our conditions both dectin-1 and the control CD74 are detected. On the right is an immunoblot of precipitates from lysates to which anti-galectin-3 antibody was added. HA-tagged dectin-1 is detected after immunoprecipitation with the anti-galectin-3 antibody and the levels increase after zymosan presentation (HA tag). Constant levels of galectin-3 are observed in all samples regardless whether the cells were incubated with zymosan or not (galectin-3). Tagged-CD74-expressing cells were used as a control.

**(b)** HEK 293T cells stably expressing tagged dectin-1 were transfected with a plasmid encoding 3xFLAG-tagged galectin-3. Transfectants were sortagged to label surface-exposed dectin-1 with a biotinylated probe and lysed with 0.5% NP-40. The immunoblot of the cell lysates shows detectable levels of dectin-1, galectin-3 and the control CD74 (Cell Lysates). A fraction of the cell extracts (500µg) was precleared and incubated with magnetic streptavidin Dynabeads overnight at 4°C. Samples were washed with 0.5% NP-40 and eluted with SDS-PAGE sample buffer. The eluted proteins were then separated by SDS-PAGE. The analysis by immunoblotting shows FLAG-tagged galectin-3 (Streptavidin pull-down). Another fraction of the cell extracts (500µg) was incubated with anti-FLAG-coated magnetic beads. Biotinylated and HA-tagged dectin-1 were detected by immunoblotting (anti-FLAG IP). HEK293T cells co-transfected with



Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

tagged CD74 and tagged galectin-3 were used as a control. dt: cells expressing tagged dectin-1; ct: control cells expressing tagged CD74.

**Figure A.7. Dectin-1 and galectin-3 modulate the TNF- $\alpha$  induction in RAW 264.7 macrophages.**

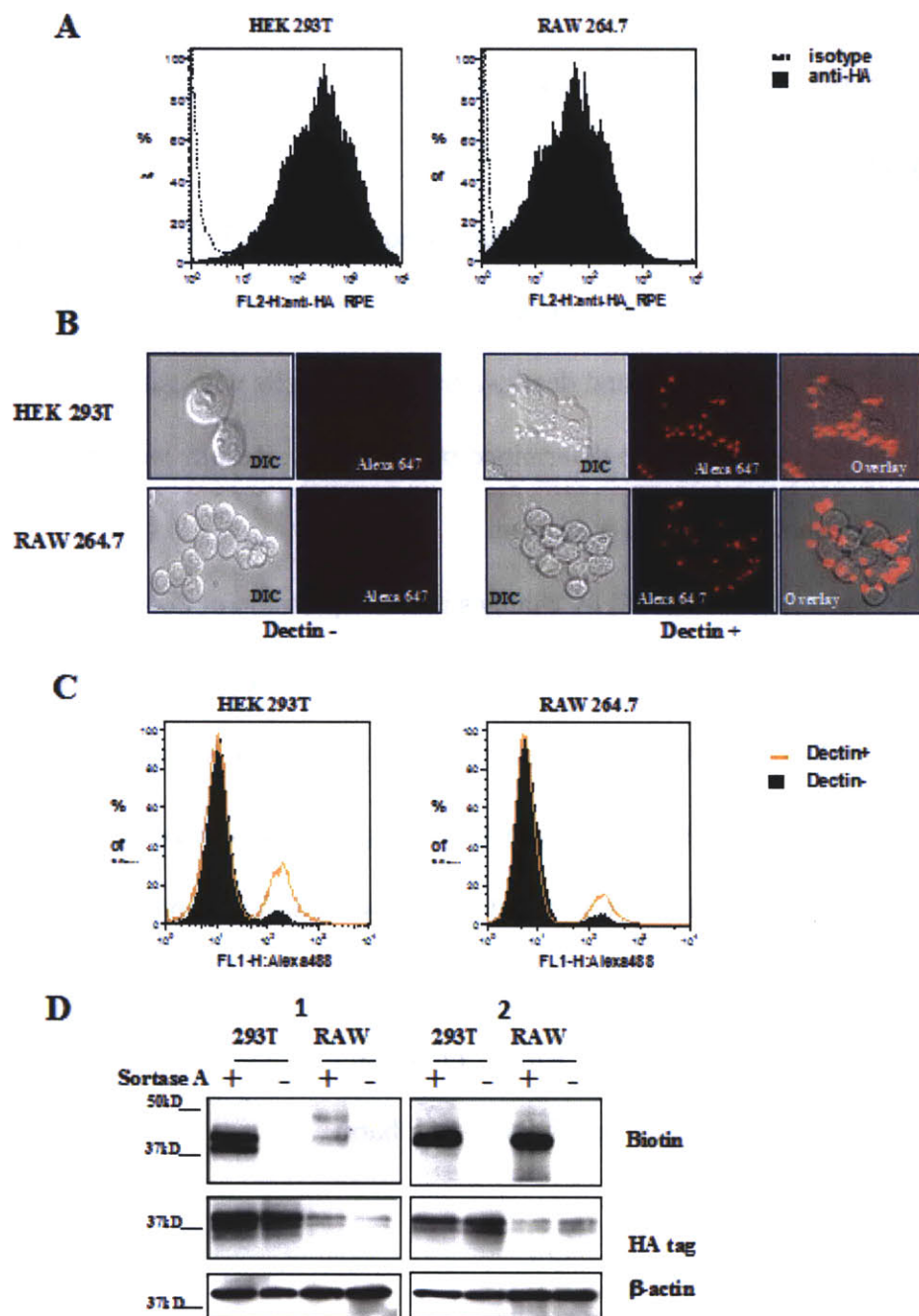
(a) Galectin-3 knockdown in dectin-1-transduced RAW 264.7 cells by siRNA transduction. Levels of mRNA from RAW 264.7 cells (RAW), dectin-1-transfected RAW 264.7 cells (RAWdect) and dectin-1-transfected cells with galectin-3 siRNA (RAWdect(shRNAGal3)) were determined by RT-qPCR. Results were normalized to GAPDH and expressed as fold induction.

(b) TNF- $\alpha$  induction is affected by changes in the expression of dectin-1 and galectin-3. Cells were incubated for 4h with UV-irradiated *S. cerevisiae* (SC) or *C. albicans* (CA) at a ratio 1:5. TNF- $\alpha$  mRNA levels were determined by RT-qPCR and normalized to GAPDH. Cells with dectin-1 expression show a significant increase of TNF- $\alpha$  mRNA levels after fungal presentation compared to cells without fungal presentation. The reduction of the levels of galectin-3 on siRNA knockdowns significantly reduced the TNF- $\alpha$  response in cells incubated with *C. albicans*. (\*p<0.05). Mean values ( $\pm$  SD) from three independent experiments are shown (error bars).

*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

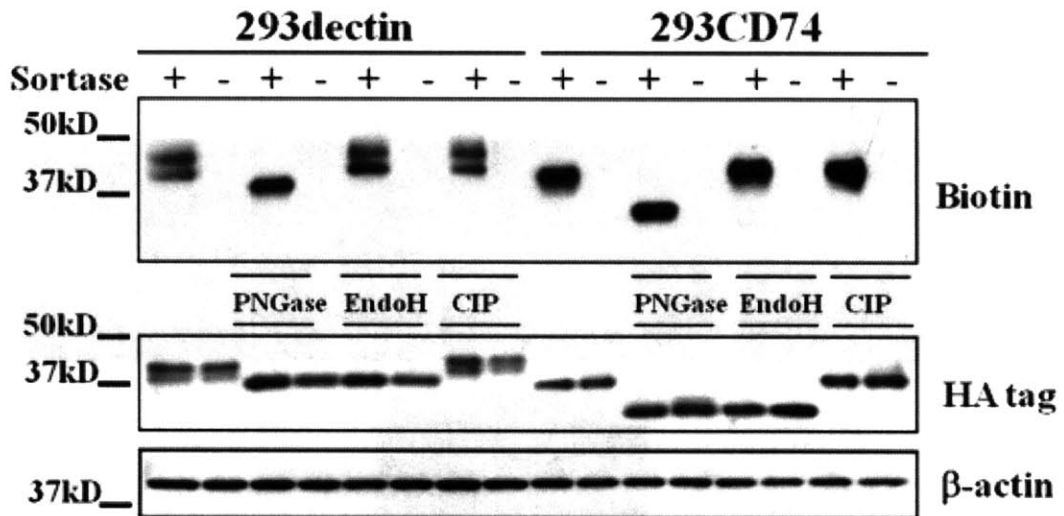
Figures

Figure A.1



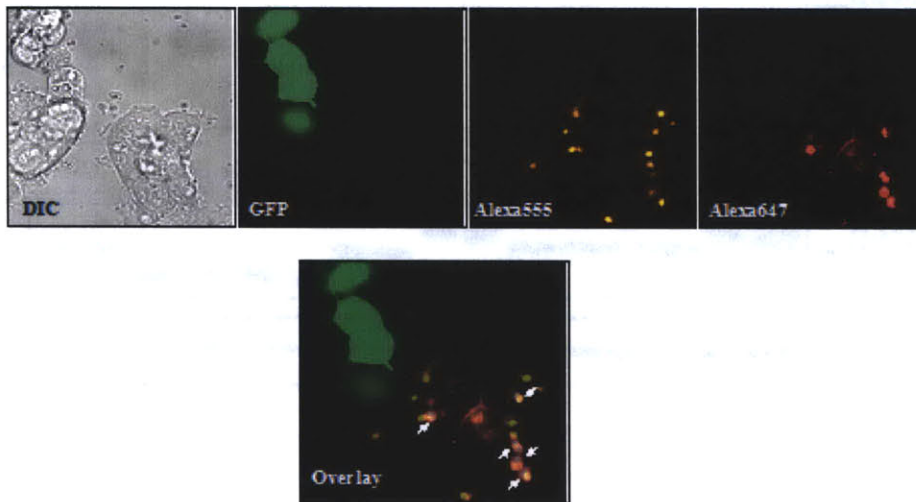
*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

Figure A.2



*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

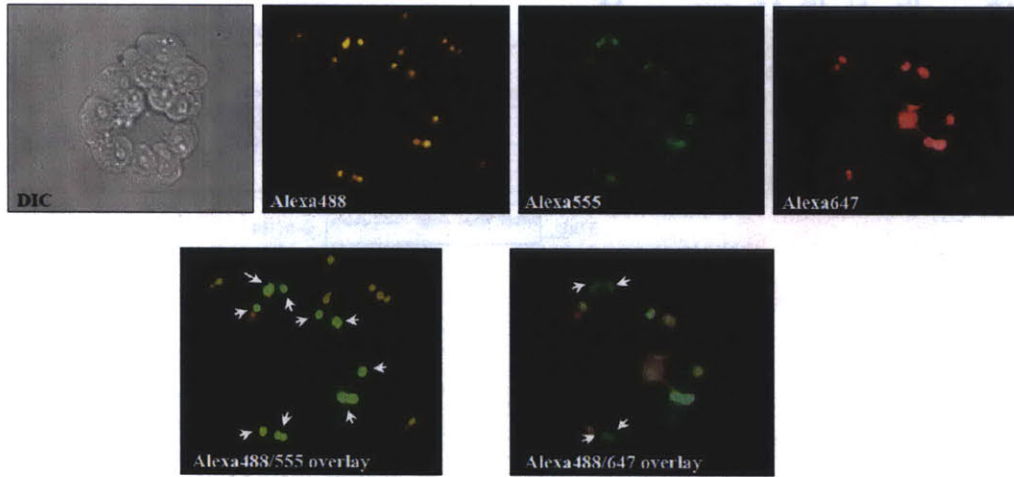
**Figure A.3**



*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

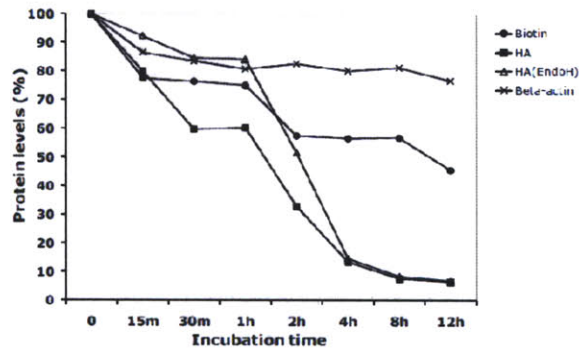
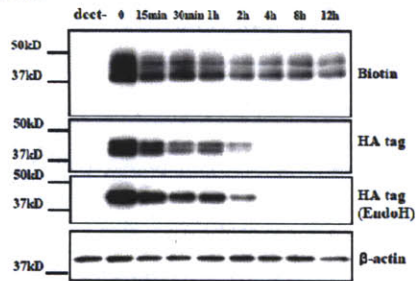
Fig. A.4

A

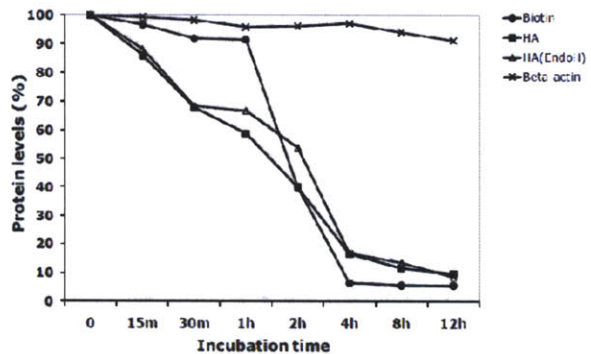
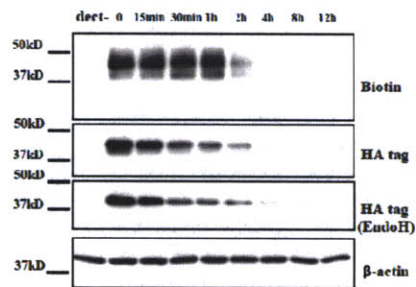


B

ZYM-

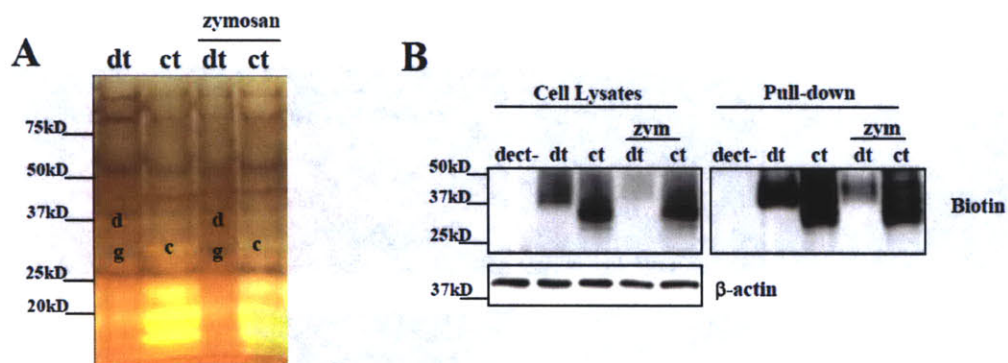


ZYM+



*Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

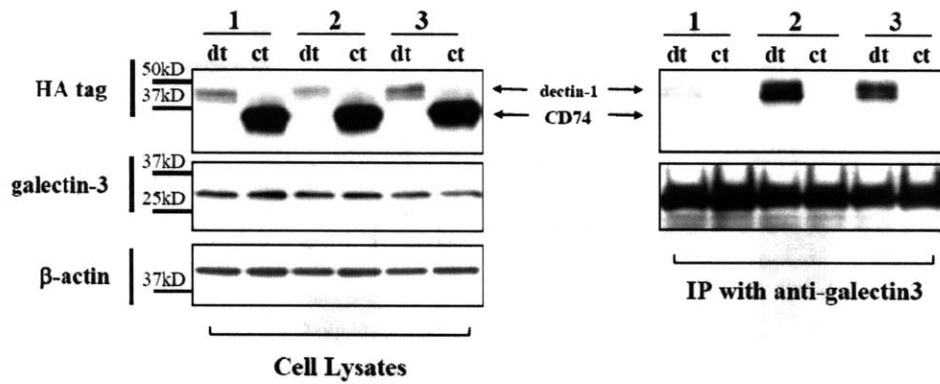
Fig. A.5



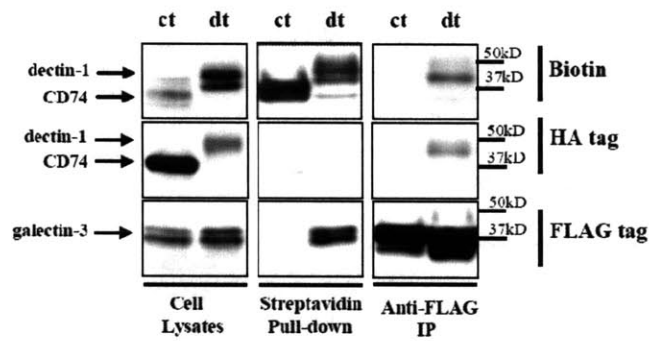
Appendix A: Sortagging reveals a functionally important association between dectin-1 and galectin-3 in macrophages

**Figure A.6**

**A**

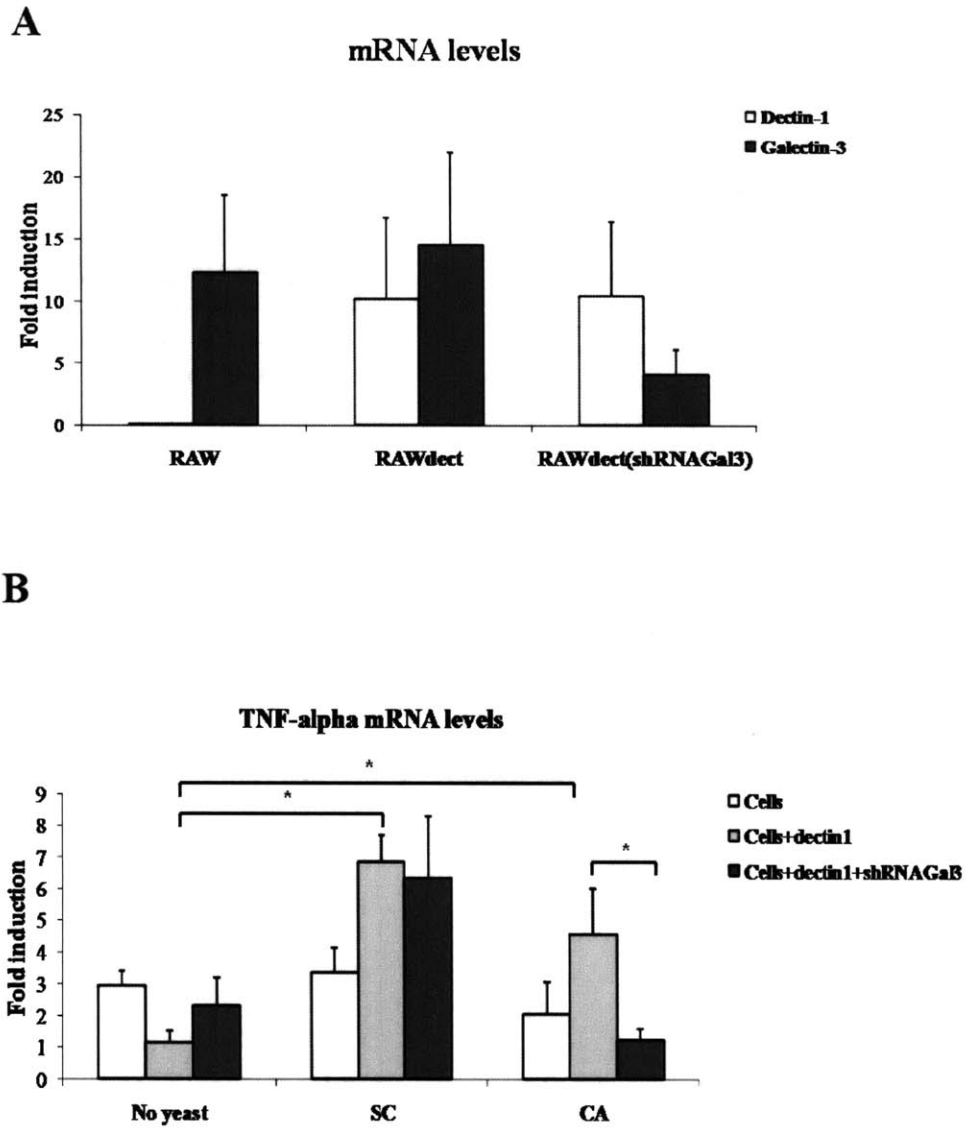


**B**



*Appendix A: Sorting reveals a functionally important association between dectin-1 and galectin-3 in macrophages*

Figure A.7





## **Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation**

Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

**Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation**

(From: M. W. Popp, J. M. Antos, H. L. Ploegh, *Curr. Protoc. Protein Sci.* 2009, Chapter 15, Unit 15.3)

**Introduction**

Creation of functional protein bioconjugates demands methods for attaching a diverse array of probes to target proteins with high specificity, under mild conditions. The sortase A transpeptidase enzyme from *Staphylococcus aureus* catalyzes the cleavage of a short 5 amino acid recognition sequence (LPXTG) with the concomitant formation of an amide linkage between an oligoglycine peptide and the target protein (**Figure B.1a**). By functionalizing the oligoglycine peptide, it is possible to incorporate reporters into target proteins in site-specific fashion. This reaction is applicable to proteins in solution and on the living cell surface. The method requires only incubation of the target protein, which has been engineered to contain a sortase recognition site either at the C terminus or within solvent accessible loops, with purified sortase enzyme and a suitably functionalized oligoglycine peptide.

**Expression and Purification of Sortase A**

Two soluble versions of the sortase A enzyme have been created for use in the transpeptidation reaction. Since sortase itself is membrane embedded, the transmembrane domain has been truncated and replaced with a hexahistidine tag to aid in purification. One version, described by the Schneewind lab <sup>1</sup>, contains an N-terminal deletion of 25 amino acids, replaced with a hexahistidine tag (cloned into pQE30, Qiagen). A second version contains an N-terminal deletion of 59 amino acids, replaced

*Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation*

by a hexahistidine tag that is separated from the body of the protein by a thrombin cleavage site (cloned into pET28a+, Novagen). These two versions have essentially identical activities<sup>2</sup>, but different mobilities in SDS-PAGE gels, a useful attribute if the substrate protein migrates at a similar molecular weight as sortase. The mobility of the  $\Delta 59$  version can be increased further by thrombin cleavage of the hexahistidine tag and linking amino acids. The following standard protein expression and purification protocol routinely yields very large quantities of sortase A (>40 mg/L of culture).

***Materials***

Luria Bertani (LB) media

Ampicillin (1000x stock = 100 mg/ml) for pQE30 derived construct

Kanamycin (1000x stock = 30 mg/ml) for pET28a(+) derived construct

Isopropyl  $\beta$ -d-thiogalactopyranoside (1 M stock)

Phosphate buffered saline (PBS)

DNaseI (10 mg/ml stock)

Lysis Buffer: 50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM imidazole, 10% glycerol

Elution Buffer: 50 mM Tris, pH 7.5, 150 mM NaCl, 500 mM imidazole, 10% glycerol

Ni-NTA agarose slurry (Qiagen)

1. Transform sortase expression plasmid into *E. coli* BL-21(DE3) and plate on selective media.

*Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation*

2. Pick a single colony and inoculate 100 ml of LB containing the appropriate antibiotic. Grow overnight as a starter culture.

3. Add 10 ml of the overnight culture into 1 L of LB with antibiotics. Monitor the OD<sub>600</sub> until ~0.7. Induce by adding IPTG to a final concentration of 1 mM and shake at 37 °C for 3 hours.

4. Harvest the bacterial pellet by centrifugation at 6000 x g for 20 minutes. Decant LB and resuspend in PBS. Repeat centrifugation to wash bacteria and decant PBS. Resuspend pellet in 20 ml of ice-cold lysis buffer with 20 µg/ml DNaseI.

*Do not add protease inhibitors to lysis buffer, as these may interfere with sortase activity.*

5. Lyse bacteria by passing through a pre-chilled French press cell twice at 1250 psi. Clarify the lysate by centrifugation at 12000 x g for 30 min.

6. Pack 2-3 ml of Ni-NTA slurry into a column and wash with 10 column volumes of lysis buffer. Apply the clarified supernatant to the column. Wash column with 50 column volumes of lysis buffer.

7. Elute sortase with 2 column volumes of elution buffer.

*Optional: Dialyze twice against 4 L of lysis buffer without imidazole to remove excess imidazole.*

Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

8. Optional: Concentrate the protein further in a centrifugal concentrator with a low molecular weight cut-off. The sortase constructs described are soluble and have been concentrated to >10 mg/ml without signs of aggregation.

**Design of Protein Substrates for Sortase-mediated Transpeptidation**

Substrates bearing the sortase recognition site (LPXTG) are readily made by standard molecular biology cloning protocols. We typically introduce glutamic acid in the X position of the recognition site as this residue is commonly found in natural substrates of sortase A<sup>3</sup>. High levels of transacylation have been achieved by placing the requisite sortase cleavage site both at the C terminus of the substrate<sup>4</sup> and in flexible loops<sup>5</sup>. For targets cleaved in flexible loops, the peptide backbone of the resulting product is interrupted and the probe is attached to the C terminus of the N-terminal cleavage product. For proteins labeled at the C terminus, we routinely add a short, flexible linker composed of Gly<sub>4</sub>Ser repeats between the body of the protein and the sortase cleavage site, although this is optional (**Figure B.1b**). Whether this linker should be included must be determined empirically, and we usually make several versions of the target protein, both omitting and varying the length of the linker region. It is important that the glycine in the minimal LPETG tag is not placed at the very C terminus; it must be in peptide linkage with at least one further C-terminal amino acid. We and others have also observed better labeling by adding an extra glycine to the C terminus of the cleavage site to yield LPETGG<sup>6-7</sup>. We routinely add a short epitope C-terminal to the LPETGG motif. This epitope is lost upon transpeptidation, providing a convenient means of monitoring the

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

efficiency of labeling by immunoblot and/or coomassie stained gel. Thus the minimal construct contains only an LPETGG fused to the C terminus of the substrate. When labeling proteins on the cell surface, it is necessary that the C terminus of the target be exposed to the extracellular culture medium; therefore, membrane proteins must be in the type II orientation, with an intracellular N terminus and an extracellular C terminus.

### **Design of Peptide Probes Compatible with Transpeptidation**

Nucleophiles compatible with sortase-mediated transpeptidation have the single structural requirement of a stretch of glycine residues with a free amino terminus (**Figure B.2a**).

Successful transpeptidation can be achieved with nucleophiles containing anywhere from 1-5 glycines, however maximum reaction rates are obtained when two or more glycines are present<sup>8-9</sup>. The versatility of the sortase-mediated labeling method lies in the remarkable tolerance of the enzyme for substituents C-terminal to the oligoglycine unit.

Synthetic nucleophiles containing 1-5 glycine residues have been decorated with a range of substituents, including fluorophores<sup>4,7</sup>, photoaffinity probes<sup>4</sup>, fatty acids<sup>10</sup>, peptide nucleic acids<sup>6</sup>, polymers<sup>11</sup>, solid supports<sup>8,11-12</sup>, or other polypeptides<sup>7,9,13</sup> allowing the site-specific ligation of these moieties to peptide and protein substrates. An example of a fluorescent triglycine nucleophile prepared by our laboratory is depicted in **Figure B.2b**.

The synthesis of oligoglycine probes is readily accomplished using standard solid phase synthesis methods. It should be noted that a variety of non-natural amino acid building blocks containing fluorophores or affinity labels are available from commercial sources and are easily integrated into manual or automated peptide syntheses. Oligoglycine

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

probes usually exhibit excellent stability, and can be stored for several months as concentrated stock solutions (100 mM - 500 mM in DMSO or water at 4 °C) with no loss of activity.

### **Labeling Protocols**

#### **Labeling Purified Proteins**

This protocol describes the labeling of proteins that have been purified prior to the transpeptidation reaction. This is useful for target proteins that can be produced in *E. coli* or in tissue culture supernatants and purified by immobilized metal affinity chromatography (see **Design of Protein Substrates for Sortase-mediated Transpeptidation section**). Progress of the reaction is monitored in accordance with the oligoglycine probe to be installed; for probes with a biotin moiety, immunoblotting suffices, while for probes containing a fluorophore, in-gel fluorescence scanning can be used. Sortase, as well as input material, can be removed in a single step from the crude reaction mixture to yield the purified transpeptidation product.

#### **Materials**

Sortase buffer (10x): 500 mM Tris, pH 7.5, 1.5 M NaCl, 100 mM CaCl<sub>2</sub>

Oligoglycine probe: 100 mM – 500 mM stock solution in DMSO or H<sub>2</sub>O

Purified target protein (Protein must not be dissolved in phosphate buffer)

Purified sortase A stock solution

Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

1. Mix ingredients in a microcentrifuge tube such that the final concentrations are as follows: 10  $\mu$ M – 50  $\mu$ M purified target protein, 150  $\mu$ M sortase A, and 1 mM oligoglycine probe. Incubate the reaction at 37 °C for 1 hour followed by analysis using SDS-PAGE and immunoblot/coomassie stain.

*The reaction conditions described above provide a suitable starting point for optimizing the labeling protocol. Slightly different reaction rates are observed for each purified target protein and consequently the reaction parameters may require optimization to achieve high levels of protein labeling. Longer reaction times are usually all that is necessary to drive the transpeptidation reaction to completion. Additional sortase may also be added to improve reaction rates. Often, the concentrations of sortase and oligoglycine probe may be reduced without sacrificing labeling efficiency. Reactions can also be incubated at room temperature if desired.*

2. If desired, His<sub>6</sub>-tagged sortase and unreacted target protein (bearing a His<sub>6</sub> tag C-terminal to the sortase cleavage site) can be removed by passing crude reaction mixtures over a Ni-NTA column in the presence of imidazole (10-30 mM) and NaCl (500 mM) to prevent nonspecific binding of proteins lacking the His<sub>6</sub> affinity tag. The column flow-through will contain the desired transpeptidation product.

### **Labeling Cell-surface Proteins in Living Cells**

In this protocol, type II membrane proteins with their C-terminus exposed to extracellular space are fused to the requisite sortase recognition motif and expressed in living cells.



*Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation*

These proteins can be labeled by introduction of sortase and the oligoglycine probe into the tissue culture media. After washing, labeled cells can be imaged directly.

***Materials***

Plasmid encoding target protein

Transfection reagent

Culture medium (the presence of 10% serum does not inhibit the reaction)

Oligoglycine probe: 100 mM – 500 mM stock solution in DMSO or H<sub>2</sub>O

Purified sortase A stock solution

PBS++ (Phosphate-buffered saline with 1 mM CaCl<sub>2</sub> and 1 mM MgCl<sub>2</sub>)

1. Transfect target cells with plasmid encoding engineered target protein according to manufacturer's directions.

*For microscopy, cells may be cultured and transfected on plastic dishes and subsequently detached and replated onto glass coverslips for labeling and analysis. Alternatively, cells can be cultured and transfected directly on chambered coverslips (Lab-Tek II Chambered Coverglass, Nunc).*

2. Incubate cells with 200 μM sortase A and 100 μM oligoglycine probe diluted in normal culture media at 37 °C in a humidified incubator for 10-30 minutes.

*As in the case of labeling purified proteins in solution, the above conditions provide an appropriate starting point for further optimization. Sortase A concentration and reaction times should be varied to achieve maximum labeling efficiency.*

Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

3. Wash cells extensively with PBS++.

*This step removes both sortase and excess probe. Extensive washing is necessary to remove unbound probe and decrease the background signal. PBS supplemented with  $MgCl_2$  and  $CaCl_2$  helps cells maintain adherence to the coverslip during washing.*

4. Add either phenol-red free media or PBS to cells and observe by microscopy.

Or

Lyse cells and load on SDS-PAGE for immunoblot or in-gel fluorescence scanning.

### **Background Information**

Site-specific incorporation of reporter molecules into proteins is a significant challenge. Common chemical labeling techniques targeting cysteine or lysine residues using maleimides or N-hydroxysuccinimidyl esters, respectively, generally lack the specificity needed for single-site labeling within a given protein, and cannot be used to label a target protein in complex mixtures such as cell lysates or living cells. Genetic fusions, while suited for single-site labeling, are mostly limited to bulky fluorescent protein tags as reporters, although other modules with enzymatic activity have been used. Thus, we and others have developed chemoenzymatic labeling techniques that exploit the exquisite

Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

specificity of enzymes to target chemical reporters to a single site within a protein bearing multiple chemically reactive side-chains.

Several alternative chemoenzymatic labeling strategies have been described and key features of each have been reviewed<sup>14-15</sup>. Chemoenzymatic methods and small molecule binding peptide sequences that allow site-specific incorporation of labels have been developed, including transglutaminase-catalyzed reactions<sup>16</sup>, acyl carrier-protein-based labeling<sup>17</sup>, O<sub>6</sub>-alkylguanine DNA alkyltransferase fusions<sup>18</sup>, dihydrofolate reductase fusions<sup>19</sup>, biotin ligase<sup>20</sup>, FAsH<sup>21</sup>, and Texas-red binding peptide<sup>22</sup>. However, several of these methods require the installation of a protein-sized module to afford selective labeling, or necessitate the insertion of recognition sequences that vary in size from 6 to 38 residues, with varying degrees of labeling selectivity. The nature of the labeling method also dictates the types of reporter molecules that can be installed, some of which require synthetic capabilities beyond the reach of most laboratories involved in biochemical or cell-biological studies. Although the sortase-mediated transpeptidation method as described is only applicable to labeling of proteins at the C terminus and solvent accessible loops, the major advantage is its simplicity. The method requires little more than sortase production and purification from *E. coli* and oligoglycine nucleophile synthesis. Construction of peptide nucleophiles by solid-phase peptide synthesis is straightforward.

Sortase A from *S. aureus* has already found numerous applications in the semi-synthesis of protein and peptide conjugates<sup>23</sup>. These include installation of fluorophores<sup>4,7</sup>,

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

photoaffinity cross-linking agents<sup>4</sup>, peptide nucleic acids with unique cell-penetrating properties<sup>6</sup>, and carbohydrates<sup>24</sup>. Cell surface labeling provides a new tool for studying the trafficking and behavior of a particular protein in live cells by microscopy.

The biochemical details of the transpeptidation reaction catalyzed by sortase have been studied<sup>25</sup>. Sortase makes hydrophobic contacts between residues in its  $\beta 6/\beta 7$  loop and the LPXTG motif on target proteins<sup>26-27</sup>. This positions a key cysteine residue (Cys-184) to attack the threonine-glycine amide bond in the LPXTG motif, generating a thioacyl intermediate and releasing all material C-terminal to the threonine (**Fig. 1**). The acyl-enzyme intermediate is then resolved by nucleophilic attack by the N-terminus of an oligoglycine probe that has been functionalized. The overall reaction mechanism is conceptually similar to cleavage by a cysteine protease, but instead of water attacking the acyl-enzyme intermediate to yield a hydrolysis product, sortase accepts the N-terminus of oligoglycine to yield a transpeptidation product.

### **Critical Parameters**

It is critical that the LPETG cleavage site is accessible to sortase. Where possible, it is helpful to view the crystal structure of the target protein; examination of the C terminus or target loop(s) will reveal whether they are likely to be solvent accessible. Where sortase may not be able to access the cleavage site due to steric interference, it may be helpful to extend the C terminus with a flexible linker or to enlarge the target loop (see substrate design section).

*Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation*

The activity of sortase requires calcium in the reaction buffers. For labeling of purified proteins, 10 mM calcium chloride is included, and for labeling of cell surface proteins, the calcium present in normal culture medium suffices for activity.

**Troubleshooting**

Low labeling efficiency of a target protein can usually be overcome simply by adjusting the reaction conditions outlined in the standard protocol. Longer incubation times, combined with varying the enzyme to target protein ratio can be helpful.

Mixing of the sortase reaction buffer with proteins in phosphate buffers should be avoided, since calcium phosphate will precipitate and inhibit labeling.

When reaction products are analyzed by SDS-PAGE, it is often helpful to remove unbound probe from the gel to lower background. For fluorescent probes, this is especially helpful before fluorescent in-gel scanning and is readily accomplished by washing several times in coomassie destain solution (30% ethanol, 10% acetic acid), and then incubating in PBS to restore neutral pH. To further limit background staining, it is helpful to titrate the amount of oligoglycine peptide to determine the minimum probe concentration that is required to achieve efficient labeling.

For cell surface labeling of proteins, repeated washing is essential to remove excess probe and achieve adequate signal to noise ratios for microscopy.

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

### **Anticipated Results**

After optimization of the basic protocol, it is often possible to achieve complete conversion of purified proteins to the transpeptidation product, yielding a homogeneously labeled target protein, as assessed by SDS-PAGE. For cell-surface labeling, signal is usually detected within 10 minutes of incubation and increases with longer reaction times. Several cell lines have been tested; target proteins in HEK293T, MDCK, CHO, and HeLa cells have all been labeled using sortase-mediated transpeptidation <sup>4,7</sup>.

### **Time Considerations**

Production and purification of the sortase A enzyme can usually be accomplished in 2 days. Probe synthesis times vary, but generally can be accomplished within a week.

With the reaction components in hand, the labeling reaction is rapid; 10 minutes suffices for detection of cell surface proteins and a few hours for complete conversion of certain purified proteins. Labeling conditions vary with the target protein, however, and must be optimized to the particular application.

Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

**References**

1. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci U S A* **96**, 12424-12429 (1999).
2. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O. & Clubb, R.T. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* **98**, 6056-6061 (2001).
3. Boekhorst, J., de Been, M.W., Kleerebezem, M. & Siezen, R.J. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J Bacteriol* **187**, 4928-4934 (2005).
4. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707-708 (2007).
5. Popp, M.W., Artavanis-Tsakonas, K. & Ploegh, H.L. Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations. *J Biol Chem* (2008).
6. Pritz, S. et al. Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. *J. Org. Chem.* **72**, 3909-3912 (2007).
7. Tanaka, T., Yamamoto, T., Tsukiji, S. & Nagamune, T. Site-specific protein modification on living cells catalyzed by Sortase. *Chembiochem* **9**, 802-807 (2008).
8. Chan, L. et al. Covalent Attachment of Proteins to Solid Supports and Surfaces via Sortase-Mediated Ligation. *PLoS ONE* **2**, e1164 (2007).
9. Mao, H., Hart, S.A., Schink, A. & Pollok, B.A. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **126**, 2670-2671 (2004).
10. Antos, J.M., Miller, G.M., Grotenbreg, G.M. & Ploegh, H.L. Lipid Modification of Proteins through Sortase-Catalyzed Transpeptidation. *J Am Chem Soc* (2008).
11. Parthasarathy, R., Subramanian, S. & Boder, E.T. Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation. *Bioconjug. Chem.* **18**, 469-476 (2007).
12. Clow, F., Fraser, J.D. & Proft, T. Immobilization of proteins to biacore sensor chips using *Staphylococcus aureus* sortase A. *Biotechnol. Lett.* **30**, 1603-1607 (2008).
13. Pritz, S. et al. Synthesis of protein mimics with nonlinear backbone topology by a combined recombinant, enzymatic, and chemical synthesis strategy. *Angew. Chem. Int. Ed. Engl.* **47**, 3642-3645 (2008).
14. Chen, I. & Ting, A.Y. Site-specific labeling of proteins with small molecules in live cells. *Curr Opin Biotechnol* **16**, 35-40 (2005).
15. Foley, T.L. & Burkart, M.D. Site-specific protein modification: advances and applications. *Curr Opin Chem Biol* **11**, 12-19 (2007).
16. Lin, C.W. & Ting, A.Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J Am Chem Soc* **128**, 4542-4543 (2006).

Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

17. George, N., Pick, H., Vogel, H., Johnsson, N. & Johnsson, K. Specific labeling of cell surface proteins with chemically diverse compounds. *J Am Chem Soc* **126**, 8896-8897 (2004).
18. Keppler, A. et al. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **21**, 86-89 (2003).
19. Miller, L.W., Sable, J., Goelet, P., Sheetz, M.P. & Cornish, V.W. Methotrexate conjugates: a molecular in vivo protein tag. *Angew Chem Int Ed Engl* **43**, 1672-1675 (2004).
20. Chen, I., Howarth, M., Lin, W. & Ting, A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2**, 99-104 (2005).
21. Griffin, B.A., Adams, S.R. & Tsien, R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272 (1998).
22. Marks, K.M., Rosinov, M. & Nolan, G.P. In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem Biol* **11**, 347-356 (2004).
23. Pritz, S. et al. Synthesis of protein mimics with nonlinear backbone topology by a combined recombinant, enzymatic, and chemical synthesis strategy. *Angew Chem Int Ed Engl* **47**, 3642-3645 (2008).
24. Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V.K. & Roy, R.P. Peptide-sugar ligation catalyzed by transpeptidase sortase: a facile approach to neoglycoconjugate synthesis. *J Am Chem Soc* **130**, 2132-2133 (2008).
25. Marraffini, L.A., Dedent, A.C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* **70**, 192-221 (2006).
26. Bentley, M.L., Lamb, E.C. & McCafferty, D.G. Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J Biol Chem* **283**, 14762-14771 (2008).
27. Bentley, M.L., Gaweska, H., Kielec, J.M. & McCafferty, D.G. Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J Biol Chem* **282**, 6571-6581 (2007).



Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated  
Transpeptidation

**Figure Legends**

**Figure B.1. Site-specific labeling of target proteins by sortase-mediated  
transpeptidation**

(a) Sortase-mediated transpeptidation mechanism. Sortase A recognizes substrates with an LPXTG motif, cleaving the peptide bond between the threonine and glycine (top) and resulting in a thioacyl intermediate (middle). A modified oligoglycine nucleophile then attacks the thioacyl intermediate to yield the transpeptidation product with the probe in amide linkage to the target protein (bottom).

(b) Substrate design. Substrate proteins typically have the sortase cleavage site (LPETGG) separated from the body of the protein by an optional Gly<sub>4</sub>Ser linker of variable length. Placement of an epitope tag (His<sub>6</sub>, HA-tag, BirA acceptor peptide) C-terminal to the LPETGG motif allows a convenient means of purification and assessing the reaction progress, as this tag is lost upon transpeptidation.

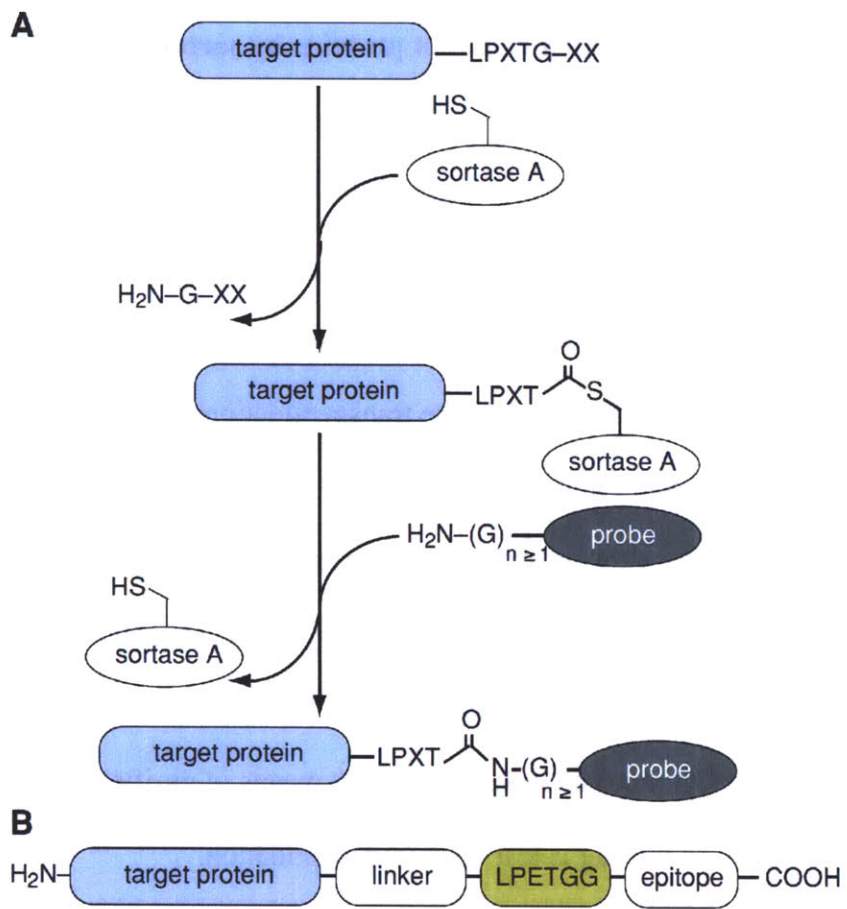
**Figure B.2. Peptide probes compatible with sortase-mediated transpeptidation.**

(a) General structure of oligoglycine based nucleophiles used in sortase-mediated transpeptidation.

(b) Example of a fluorescent probe, consisting of an oligoglycine peptide unit followed by a lysine to which fluorescein is appended.

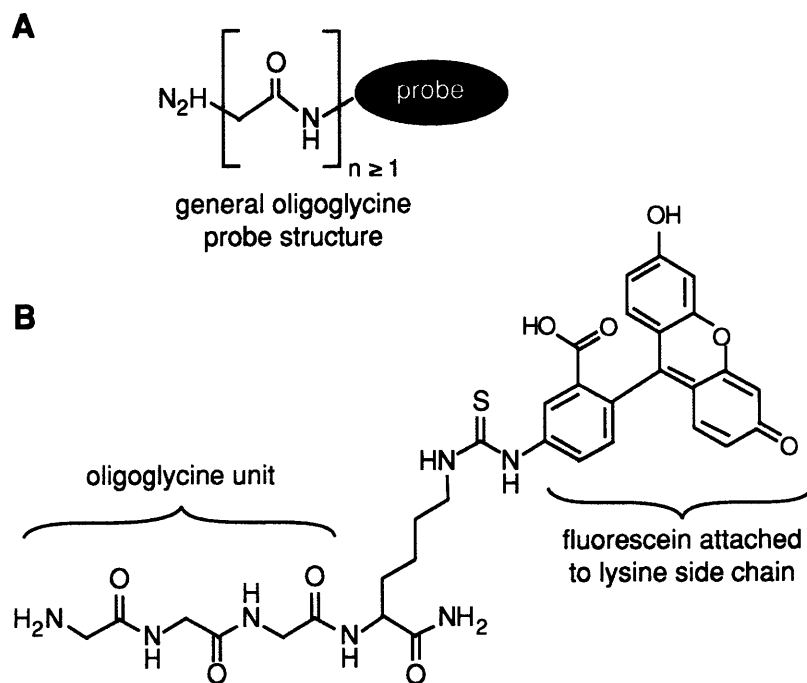
*Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation*

**Figure B.1**



Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

Figure B.2



## Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation

Maximilian Wei-Lin Popp,<sup>1,2</sup> John M. Antos,<sup>1</sup> and Hidde L. Ploegh<sup>1,2</sup>

<sup>1</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts

<sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts

### ABSTRACT

Creation of functional protein bioconjugates demands methods for attaching a diverse array of probes to target proteins with high specificity, under mild conditions. The sortase A transpeptidase enzyme from *Staphylococcus aureus* catalyzes the cleavage of a short 5-aa recognition sequence (LPXTG) with the concomitant formation of an amide linkage between an oligoglycine peptide and the target protein. By functionalizing the oligoglycine peptide, it is possible to incorporate reporters into target proteins in a site-specific fashion. This reaction is applicable to proteins in solution and on the living cell surface. The method described in this unit only requires incubation of the target protein, which has been engineered to contain a sortase recognition site either at the C terminus or within solvent-accessible loops, with a purified sortase enzyme and a suitably functionalized oligoglycine peptide. *Curr. Protoc. Protein Sci.* 56:15.3.1-15.3.9. © 2009 by John Wiley & Sons, Inc.

Keywords: sortase • transpeptidation • site-specific labeling • chemoenzymatic labeling

### INTRODUCTION

Creation of functional protein bioconjugates demands methods for attaching a diverse array of probes to target proteins with high specificity, under mild conditions. The sortase A transpeptidase enzyme from *Staphylococcus aureus* catalyzes the cleavage of a short 5-aa recognition sequence (LPXTG) with the concomitant formation of an amide linkage between an oligoglycine peptide and the target protein (Fig. 15.3.1A). By functionalizing the oligoglycine peptide, it is possible to incorporate reporters into target proteins in a site-specific fashion. This reaction is applicable to proteins in solution and on the living cell surface. The method only requires incubation of the target protein, which has been engineered to contain a sortase recognition site either at the C terminus or within solvent-accessible loops, with a purified sortase enzyme and a suitably functionalized oligoglycine peptide.

### STRATEGIC PLANNING

#### Design of Protein Substrates for Sortase-Mediated Transpeptidation

Substrates bearing the sortase recognition site (LPXTG) are readily made using standard molecular biology cloning protocols. It is convenient to introduce glutamic acid in the X position of the recognition site, as this residue is commonly found in natural substrates of sortase A (Boekhorst et al., 2005). High levels of transacylation have been achieved by placing the requisite sortase cleavage site both at the C terminus of the substrate (Popp et al., 2007) and in flexible loops (Popp et al., 2008). For targets cleaved in flexible loops, the peptide backbone of the resulting product is interrupted and the probe is attached to the C terminus of the N-terminal cleavage product. For proteins labeled at the C terminus, the authors routinely add a short, flexible linker composed of Gly<sub>4</sub>Ser repeats between the body of the protein and the sortase cleavage site, although this is optional (Fig. 15.3.1B). Whether this linker should be included must be determined empirically,

*Current Protocols in Protein Science* 15.3.1-15.3.9, April 2009  
Published online April 2009 in Wiley InterScience (www.interscience.wiley.com).  
DOI: 10.1002/0471140864.ps1503s56  
Copyright © 2009 John Wiley & Sons, Inc.

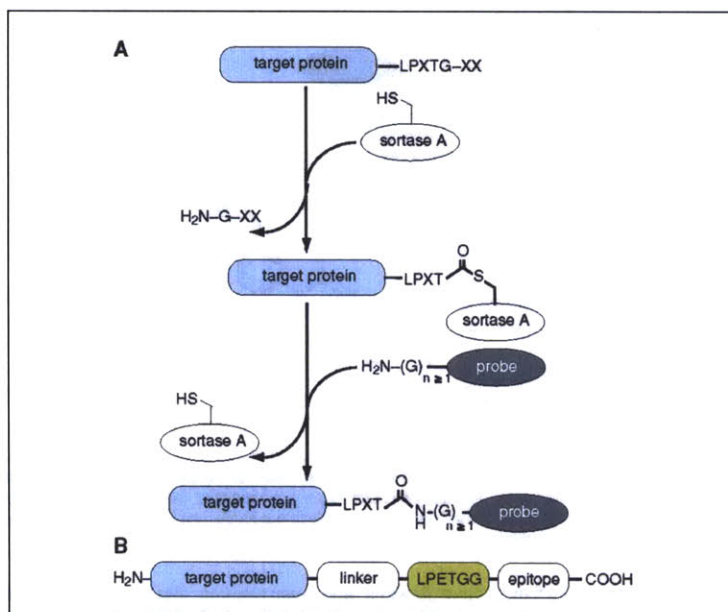
UNIT 15.3

Chemical  
Modification of  
Proteins

15.3.1

Supplement 56

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation



**Figure 15.3.1** Site-specific labeling of target proteins by sortase-mediated transpeptidation. (A) Sortase-mediated transpeptidation mechanism. Sortase A recognizes substrates with an LPXTG motif, cleaving the peptide bond between the threonine and glycine (top) and resulting in a thioacyl intermediate (middle). A modified oligoglycine nucleophile then attacks the thioacyl intermediate to yield the transpeptidation product with the probe in amide linkage to the target protein (bottom). (B) Substrate design. Substrate proteins typically have the sortase cleavage site (LPETGG) separated from the body of the protein by an optional Gly<sub>3</sub>Ser linker of variable length. Placement of an epitope tag (His<sub>6</sub>, HA-tag, BirA acceptor peptide) C-terminal to the LPETGG motif allows a convenient means of purification and assessing the reaction progress, as this tag is lost upon transpeptidation.

and it is recommended to make several versions of the target protein, both omitting and varying the length of the linker region. It is important that the glycine in the minimal LPETG tag is not placed at the very C terminus; it must be in peptide linkage with at least one further C-terminal amino acid. Better labeling is achieved by adding an extra glycine to the C terminus of the cleavage site to yield LPETGG (Pritz et al., 2007; Tanaka et al., 2008). The authors also routinely add a short epitope C-terminal to the LPETGG motif. This epitope is lost upon transpeptidation, providing a convenient means of monitoring the efficiency of labeling by immunoblotting and/or a Coomassie stained gel. Thus, the minimal construct contains only an LPETGG fused to the C terminus of the substrate. When labeling proteins on the cell surface, it is necessary that the C terminus of the target be exposed to the extracellular culture medium; therefore, membrane proteins must be in the type II orientation, with an intracellular N terminus and an extracellular C terminus.

### Design of Peptide Probes Compatible with Transpeptidation

Nucleophiles compatible with sortase-mediated transpeptidation have the single structural requirement of a stretch of glycine residues with a free amino terminus (Fig. 15.3.2A). Successful transpeptidation can be achieved with nucleophiles containing

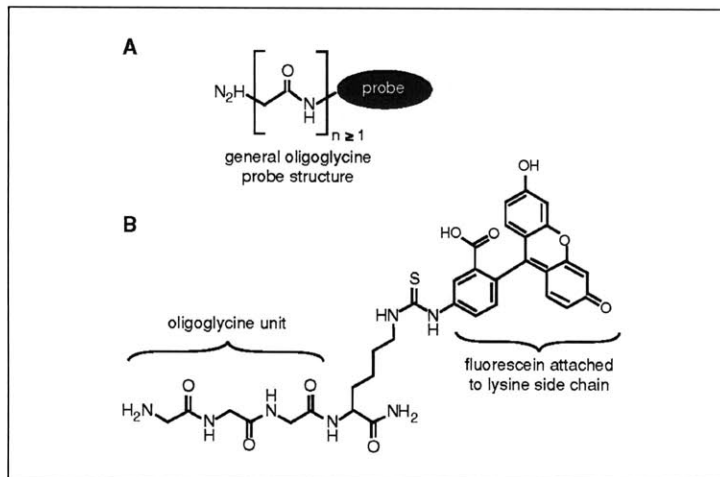
Labeling via  
Sortase-mediated  
Transpeptidation

### 15.3.2

Supplement 56

Current Protocols in Protein Science

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation



**Figure 15.3.2** Peptide probes compatible with sortase-mediated transpeptidation. (A) General structure of oligoglycine based nucleophiles used in sortase-mediated transpeptidation. (B) Example of a fluorescent probe, consisting of an oligoglycine peptide unit followed by a lysine to which fluorescein is appended.

anywhere from one to five glycines; however, maximum reaction rates are obtained when two or more glycines are present (Mao et al., 2004; Chan et al., 2007). The versatility of the sortase-mediated labeling method lies in the remarkable tolerance of the enzyme for the substituent C-terminal to the oligoglycine unit. Synthetic nucleophiles containing one to five glycine residues have been decorated with a range of substituents, including fluorophores (Popp et al., 2007; Tanaka et al., 2008), photoaffinity probes (Popp et al., 2007), fatty acids (Antos et al., 2008), peptide nucleic acids (Pritz et al., 2007), polymers (Parthasarathy et al., 2007), solid supports (Chan et al., 2007; Parthasarathy et al., 2007; Clow et al., 2008), or other polypeptides (Mao et al., 2004; Pritz et al., 2008; Tanaka et al., 2008) allowing the site-specific ligation of these moieties to peptide and protein substrates. An example of a fluorescent triglycine nucleophile is depicted in Figure 15.3.2B.

The synthesis of oligoglycine probes is readily accomplished using standard solid-phase synthesis methods (see Chapter 18). It should be noted that a variety of non-natural amino acid building blocks containing fluorophores or affinity labels are available from commercial sources (Novabiochem, AnaSpec, Advanced Chemtech) and are easily integrated into manual or automated peptide syntheses. Oligoglycine probes usually exhibit excellent stability and can be stored for several months as concentrated stock solutions (100 mM to 500 mM in DMSO or water at 4°C) with no loss of activity.

### SITE-SPECIFIC PROTEIN LABELING VIA SORTASE-MEDIATED TRANSPEPTIDATION

This protocol describes the labeling of proteins that have been purified prior to the transpeptidation reaction. This is useful for target proteins that can be produced in *E. coli* or in tissue culture supernatants and purified by immobilized metal affinity chromatography (see Strategic Planning). Progress of the reaction is monitored in accordance with the oligoglycine probe to be installed; immunoblotting can be used for probes with a biotin moiety, while in-gel fluorescence scanning is convenient for probes containing a

### BASIC PROTOCOL

### Chemical Modification of Proteins

### 15.3.3

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

fluorophore. Sortase, as well as input material, can be removed from the crude reaction mixture in a single step to yield the purified transpeptidation product.

### **Materials**

Purified target protein (protein must not be dissolved in phosphate buffer)  
Purified sortase A stock solution (see Support Protocol)  
100 mM to 500 mM oligoglycine probe stock solution in DMSO or H<sub>2</sub>O  
Ni-NTA column, optional  
10 to 30 mM imidazole, optional  
500 mM NaCl, optional  
10× sortase buffer: 500 mM Tris-Cl, pH 7.5 (APPENDIX 2E), 1.5 M NaCl, 100 mM CaCl<sub>2</sub>

1.5-ml microcentrifuge tubes  
37°C water bath

Additional reagents and equipment for SDS-PAGE (UNIT 10.1), immunoblot (UNITS 10.7 & 10.10), and Coomassie stain (UNIT 10.5)

1. Mix the following in a 1.5-ml microcentrifuge tube such that the final concentrations are as follows:

10 μM to 50 μM purified target protein  
150 μM sortase A  
1 mM oligoglycine probe  
1× sortase buffer.

Incubate the reaction 1 hr at 37°C followed by analysis using SDS-PAGE (UNIT 10.1) and immunoblotting (UNITS 10.7 & 10.10) or Coomassie staining (UNIT 10.5).

*The reaction conditions described above provide a suitable starting point for optimizing the labeling protocol. Slightly different reaction rates are observed for each purified target protein and consequently the reaction parameters may require optimization to achieve high levels of protein labeling. Longer reaction times are usually all that is necessary to drive the transpeptidation reaction to completion. Additional sortase may also be added to improve reaction rates. Often, the concentrations of sortase and oligoglycine probe may be reduced without sacrificing labeling efficiency. Reactions can also be incubated for a longer period of time (1 to 2 hr) at room temperature if desired.*

2. (Optional) Remove His<sub>6</sub>-tagged sortase and unreacted target protein (bearing a His<sub>6</sub> tag C-terminal to the sortase cleavage site) by passing crude reaction mixtures over a Ni-NTA column in the presence of imidazole (10 to 30 mM) and NaCl (500 mM) to prevent nonspecific binding of proteins lacking the His<sub>6</sub> affinity tag.

*The column flow-through will contain the desired transpeptidation product.*

### **ALTERNATE PROTOCOL**

### **LABELING CELL-SURFACE PROTEINS IN LIVING CELLS VIA SORTASE-MEDIATED TRANSPEPTIDATION**

In this protocol, type II membrane proteins with their C-terminus exposed to extracellular space are fused to the requisite sortase recognition motif and expressed in living cells. These proteins can be labeled by introduction of sortase and the oligoglycine probe into the tissue culture medium. After washing, labeled cells can be imaged directly.

### **Materials**

Target cells  
Plasmid encoding target protein  
Transfection reagent (Lipofectamine, Invitrogen; Trans IT, Minus; Fugene 6, Roche)

Labeling via  
Sortase-mediated  
Transpeptidation

15.3.4

Supplement 56

Current Protocols in Protein Science

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

Purified sortase A stock solution (see Support Protocol)  
100 mM to 500 mM oligoglycine probe stock solution in DMSO or H<sub>2</sub>O  
Culture medium (presence of 10% serum does not inhibit the sortase reaction)  
Phosphate-buffered saline with 1 mM CaCl<sub>2</sub> and 1 mM MgCl<sub>2</sub> (PBS++)  
Plastic culture dishes or chambered coverslips (Lab-Tek II Chambered Coverglass, Nunc)  
37°C humidified incubator  
Additional reagents and equipment for SDS-PAGE (UNIT 10.1) and immunoblot (UNITS 10.7 & 10.10)

1. Transfect target cells with plasmid encoding engineered target protein according to transfection reagent manufacturer's directions.

*For microscopy, cells may be cultured and transfected on plastic dishes and subsequently detached and replated onto glass coverslips for labeling and analysis. Alternatively, cells can be cultured and transfected directly on chambered coverslips (Lab-Tek II Chambered Coverglass, Nunc).*

2. Incubate cells with 200 μM sortase A and 100 μM oligoglycine probe diluted in normal culture medium 10 to 30 min in a 37°C humidified incubator.

*As in the case of labeling purified proteins in solution (see Basic Protocol), the above conditions provide an appropriate starting point for further optimization. Sortase A concentration and reaction times should be varied to achieve maximum labeling efficiency.*

3. Wash cells at least three times with 1 ml of PBS++.

*This step removes both sortase and excess probe. Extensive washing is necessary to remove unbound probe and decrease the background signal. PBS supplemented with MgCl<sub>2</sub> and CaCl<sub>2</sub> helps cells maintain adherence to the coverslip during washing.*

4. Add either phenol-red free medium or PBS to cells and observe by microscopy. Alternatively, lyse cells and load on SDS-PAGE (UNIT 10.1) for immunoblot (UNITS 10.7 & 10.10) or in-gel fluorescence scanning.

### EXPRESSION AND PURIFICATION OF SORTASE A

Two soluble versions of the sortase A enzyme have been created for use in the transpeptidation reaction. Since sortase itself is membrane embedded, the transmembrane domain has been truncated and replaced with a hexahistidine tag to aid in purification. One version (Ton-That et al., 1999) contains an N-terminal deletion of 25 amino acids, replaced with a hexahistidine tag (cloned into pQE30, Qiagen). A second version contains an N-terminal deletion of 59 amino acids, replaced by a hexahistidine tag that is separated from the body of the protein by a thrombin cleavage site (cloned into pET28a+, Novagen). These two versions have essentially identical activities (Ilangovan et al., 2001), but different mobilities in SDS-PAGE gels, a useful attribute if the substrate protein migrates at a similar molecular weight as sortase. The mobility of the Δ59 version can be increased further by thrombin cleavage of the hexahistidine tag and linking amino acids. The following standard protein expression and purification protocol routinely yields very large quantities of sortase A (>40 mg/liter of culture).

#### Materials

Sortase expression plasmid: pQE30 (Qiagen) or pET28a+ (Novagen)  
*E. coli* BL-21(DE3)  
Luria Bertani (LB) medium with and without appropriate antibiotic:  
Ampicillin (1000× stock = 100 mg/ml) for pQE30-derived construct  
Kanamycin (1000× stock = 30 mg/ml) for pET28a(+)-derived construct  
1 M isopropyl β-D-thiogalactopyranoside (IPTG)

#### SUPPORT PROTOCOL

Chemical  
Modification of  
Proteins

#### 15.3.5



## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

Phosphate buffered saline (PBS; *APPENDIX 2E*)  
Lysis buffer: 50 mM Tris-Cl, pH 7.5 (*APPENDIX 2E*), 150 mM NaCl, 10 mM imidazole, 10% glycerol  
10 mg/ml DNaseI; *APPENDIX 2E*  
Ni-NTA agarose slurry (Qiagen)  
Elution buffer: 50 mM Tris-Cl, pH 7.5 (*APPENDIX 2E*), 150 mM NaCl, 500 mM imidazole, 10% glycerol

Culture plates and tubes  
37°C incubator with shaking  
Spectrophotometer  
French press, pre-chilled  
1.5 × 12-cm column (Biorad)  
Centrifugal spin concentrator with low-molecular-weight (<10 kD) cutoff optional (Millipore)

1. Transform sortase expression plasmid into *E. coli* BL-21(DE3) and plate on selective medium.
2. Pick a single colony and inoculate 100 ml of LB containing the appropriate antibiotic. Grow overnight as a starter culture.
3. Add 10 ml of the overnight culture into 1 liter of LB with antibiotics. Monitor the OD<sub>600</sub> until ~0.7. Induce by adding IPTG to a final concentration of 1 mM and shake 3 hr at 37°C.
4. Harvest the bacterial pellet by centrifuging 20 min at 6000 × g, 4°C. Decant LB and resuspend in 50 ml PBS. Repeat centrifugation to wash bacteria and decant PBS. Resuspend pellet in 20 ml of ice-cold lysis buffer with 20 µg/ml DNase I.  
*Do not add protease inhibitors to lysis buffer as these may interfere with sortase activity.*
5. Lyse bacteria by passing through a pre-chilled French press cell two times at 1250 psi. Clarify the lysate by centrifuging 30 min at 12,000 × g, 4°C.
6. Pack 2 to 3 ml of Ni-NTA agarose slurry into a column and wash with 10 column volumes of lysis buffer. Apply the clarified supernatant to the column. Wash column with 50 column volumes of lysis buffer.
7. Elute sortase with 2 column volumes of elution buffer.  
*(Optional) Dialyze two times against 4 liters of lysis buffer without imidazole to remove excess imidazole.*
8. (Optional) Concentrate the protein further in a centrifugal concentrator with a low-molecular-weight cutoff (<10 kDa).

*The sortase constructs described are soluble and have been concentrated to >10 mg/ml without signs of aggregation.*

### COMMENTARY

#### Background Information

Site-specific incorporation of reporter molecules into proteins is a significant challenge. Common chemical labeling techniques targeting cysteine or lysine residues using maleimides or *N*-hydroxysuccinimidyl esters, respectively, generally lack the specificity needed for single-site labeling within a given protein, and cannot be used to label a tar-

get protein in complex mixtures such as cell lysates or living cells. Genetic fusions, while suited for single-site labeling, are mostly limited to bulky fluorescent protein tags as reporters, although other modules with enzymatic activity have been used. Thus, chemoenzymatic labeling techniques have been developed to exploit the exquisite specificity of enzymes to target chemical reporters to a

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

single site within a protein bearing multiple chemically reactive side-chains.

Several alternative chemoenzymatic labeling strategies have been described and key features of each have been reviewed (Chen and Ting, 2005; Foley and Burkart, 2007). Chemoenzymatic methods and small molecule-binding peptide sequences that allow site-specific incorporation of labels have been developed, including transglutaminase-catalyzed reactions (Lin and Ting, 2006), acyl carrier-protein-based labeling (George et al., 2004), *O*<sub>6</sub>-alkylguanine DNA alkyltransferase fusions (Keppler et al., 2003), dihydrofolate reductase fusions (Miller et al., 2004), biotin ligase (Chen et al., 2005), FLAsH (Griffin et al., 1998), and Texas-red binding peptide (Marks et al., 2004). However, several of these methods require the installation of a protein-sized module to afford selective labeling, or necessitate the insertion of recognition sequences that vary in size from 6 to 38 residues, with varying degrees of labeling selectivity. The nature of the labeling method also dictates the types of reporter molecules that can be installed, some of which require synthetic capabilities beyond the reach of most laboratories involved in biochemical or cell-biological studies. Although the sortase-mediated transpeptidation method as described in this unit is only applicable to labeling of proteins at the C terminus and solvent-accessible loops, the major advantage is its simplicity. The method requires little more than sortase production and purification from *E. coli* and oligoglycine nucleophile synthesis. Construction of peptide nucleophiles by solid-phase peptide synthesis is straightforward.

Sortase A from *S. aureus* has already found numerous applications in the semi-synthesis of protein and peptide conjugates (Pritz et al., 2008). These include installation of fluorophores (Popp et al., 2007; Tanaka et al., 2008), photoaffinity cross-linking agents (Popp et al., 2007), peptide nucleic acids with unique cell-penetrating properties (Pritz et al., 2007), and carbohydrates (Samantaray et al., 2008). Cell surface labeling provides a new tool for studying the trafficking and behavior of a particular protein in live cells by microscopy.

The biochemical details of the transpeptidation reaction catalyzed by sortase have been studied (Marraffini et al., 2006). Sortase makes hydrophobic contacts between residues in its  $\beta 6/\beta 7$  loop and the LPXTG motif on target proteins (Bentley et al., 2007, 2008). This po-

sitions a key cysteine residue (Cys-184) to attack the threonine-glycine amide bond in the LPXTG motif, generating a thioacyl intermediate and releasing all material C-terminal to the threonine (Fig. 15.3.1). The acyl-enzyme intermediate is then resolved by nucleophilic attack by the N-terminus of an oligoglycine probe that has been functionalized. The overall reaction mechanism is conceptually similar to cleavage by a cysteine protease, but instead of water attacking the acyl-enzyme intermediate to yield a hydrolysis product, sortase accepts the N-terminus of oligoglycine to yield a transpeptidation product.

### Critical Parameters

It is critical that the LPETG cleavage site is accessible to sortase. Where possible, it is helpful to view the crystal structure of the target protein; examination of the C terminus or target loop(s) will reveal whether they are likely to be solvent accessible. In cases where sortase may not be able to access the cleavage site due to steric interference, it may be helpful to extend the C terminus with a flexible linker or to enlarge the target loop (see Strategic Planning).

The activity of sortase requires calcium in the reaction buffers. For labeling of purified proteins, 10 mM calcium chloride is included, and for labeling of cell surface proteins, the calcium present in normal culture medium suffices for activity.

### Troubleshooting

Low labeling efficiency of a target protein can usually be overcome simply by adjusting the reaction conditions outlined in the Basic Protocol. Longer incubation times, combined with varying the ratio of enzyme to target protein can be helpful.

Mixing of the sortase reaction buffer with proteins in phosphate buffers should be avoided since calcium phosphate will precipitate and inhibit labeling.

When reaction products are analyzed by SDS-PAGE, it is often helpful to remove unbound probe from the gel to reduce background. For fluorescent probes, this is especially helpful before fluorescent in-gel scanning and is readily accomplished by washing several times in Coomassie destain solution (30% ethanol, 10% acetic acid), and then incubating in PBS to restore neutral pH. To further limit background staining, it is helpful to titrate the amount of oligoglycine peptide to

Chemical  
Modification of  
Proteins

### 15.3.7

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

determine the minimum probe concentration required to achieve efficient labeling.

For cell surface labeling of proteins, repeated washing is essential to remove excess probe and achieve adequate signal-to-noise ratios for microscopy.

### Anticipated Results

After optimization of the Basic Protocol, it is often possible to achieve complete conversion of purified proteins to the transpeptidation product, yielding a homogeneously labeled target protein, as assessed by SDS-PAGE. For cell-surface labeling, signal is usually detected within 10 min of incubation and increases with longer reaction times. Several cell lines have been tested; target proteins in HEK293T, MDCK, CHO, and HeLa cells have all been labeled using sortase-mediated transpeptidation (Popp et al., 2007; Tanaka et al., 2008).

### Time Considerations

Production and purification of the sortase A enzyme can usually be accomplished in 2 days. Probe synthesis times vary, but generally can be accomplished within 1 week. With the reaction components in hand, the labeling reaction is rapid; 10 min suffices for detection of cell surface proteins and a few hours for complete conversion of certain purified proteins. Labeling conditions vary with the target protein, however, and must be optimized to the particular application.

### Literature Cited

- Antos, J.M., Miller, G.M., Grotenbreg, G.M., and Ploegh, H.L. 2008. Lipid modification of proteins through sortase-catalyzed transpeptidation. *J. Am. Chem. Soc.* epub ahead of print.
- Bentley, M.L., Gaweska, H., Kielec, J.M., and McCafferty, D.G. 2007. Engineering the substrate specificity of *Staphylococcus aureus* sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J. Biol. Chem.* 282:6571-6581.
- Bentley, M.L., Lamb, E.C., and McCafferty, D.G. 2008. Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J. Biol. Chem.* 283:14762-14771.
- Boekhorst, J., de Been, M.W., Kleerebezem, M., and Siezen, R.J. 2005. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J. Bacteriol.* 187:4928-4934.
- Chan, L., Cross, H.F., She, J.K., Cavalli, G., Martins, H.F.P., and Neylon, C. 2007. Covalent attachment of proteins to solid supports and surfaces via sortase-mediated ligation. *PLoS ONE* 211:e1164.
- Chen, I. and Ting, A.Y. 2005. Site-specific labeling of proteins with small molecules in live cells. *Curr. Opin. Biotechnol.* 16:135-40.
- 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* 22:99-104.
- Clow, F., Fraser, J.D., and Proft, T. 2008. Immobilization of proteins to biacore sensor chips using *Staphylococcus aureus* sortase A. *Biotechnol. Lett.* 309:1603-1607.
- Foley, T.L. and Burkart, M.D. 2007. Site-specific protein modification: Advances and applications. *Curr. Opin. Chem. Biol.* 11:12-19.
- George, N., Pick, H., Vogel, H., Johnsson, N., and Johnsson, K. 2004. Specific labeling of cell surface proteins with chemically diverse compounds. *J. Am. Chem. Soc.* 126:8896-8897.
- Griffin, B.A., Adams, S.R., and Tsien, R.Y. 1998. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281:269-272.
- Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R.T. 2001. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* 98:6056-6061.
- Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. 2003. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* 21:1:86-89.
- Lin, C.W. and Ting, A.Y. 2006. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* 128:4542-4543.
- Mao, H., Hart, S.A., Schink, A., and Pollok, B.A. 2004. Sortase-mediated protein ligation: A new method for protein engineering. *J. Am. Chem. Soc.* 126:2670-2671.
- Marks, K.M., Rosinov, M., and Nolan, G.P. 2004. In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem. Biol.* 113:347-356.
- Marraffini, L.A., Dedent, A.C., and Schneewind, O. 2006. Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* 70:192-221.
- Miller, L.W., Sable, J., Goelet, P., Sheetz, M.P., and Cornish, V.W. 2004. Methotrexate conjugates: A molecular in vivo protein tag. *Angew. Chem. Int. Ed. Engl.* 43:1672-1675.
- Parthasarathy, R., Subramanian, S., and Boder, E.T. 2007. Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation. *Bioconjug. Chem.* 18:2:469-476.
- 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.* 3:11:707-708.

## Appendix B: Protocol for Site-specific Protein Labeling via Sortase-mediated Transpeptidation

- Popp, M.W., Artavanis-Tsakonas, K., and Ploegh, H.L. 2008. Substrate filtering by the active-site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations. *J. Biol. Chem.* epub ahead of print.
- Pritz, S., Wolf, Y., Kraetke, O., Klose, J., Bienert, M., and Beyermann, M. 2007. Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. *J. Org. Chem.* 72:3909-3912.
- Pritz, S., Kraetke, O., Klose, A., Klose, J., Rothermund, S., Fechner, K., Bienert, M., and Beyermann, M. 2008. Synthesis of protein mimics with nonlinear backbone topology by a combined recombinant, enzymatic, and chemical synthesis strategy. *Angew. Chem. Int. Ed. Engl.* 47:3642-3645.
- Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V.K., and Roy, R.P. 2008. Peptide-sugar ligation catalyzed by transpeptidase sortase: A facile approach to neoglycoconjugate synthesis. *J. Am. Chem. Soc.* 130:2132-2133.
- Tanaka, T., Yamamoto, T., Tsukiji, S., and Nagamune, T. 2008. Site-specific protein modification on living cells catalyzed by sortase. *ChemBiochem* 95:802-807.
- Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F., and Schneewind, O. 1999. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc. Natl. Acad. Sci. U.S.A.* 96:12424-12429.