DOI: 10.1002/anie.201008267

Protein Engineering

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

Sortases 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.[1] Anchoring of proteins to the bacterial cell wall is the purview of sortase enzymes,[2–4] 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.[5–7] Other sortases from different bacterial species use the same or similar recognition sequences.[8–10] A database of sortases and their substrates can be found on the internet.[11,12] This peptide-cleaving reaction initially yields a thioacyl intermediate,[13,14] 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.[15] 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.[16–19] The details of this process differ between bacterial species,[20–24] 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.[25] 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.[26]

Sortases represent a bona fide drug target because of their central role in virulence.[27–31] Recombinant sortase lacking its transmembrane domain is readily produced in high yield.[6,32,33] For a detailed protocol for sortase production, see Ref. [34]. This has facilitated extensive structural[35–38] and biochemical[39–41] 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.[42] The biochemical details of the active site include a key histidine residue (H120) that can form a thiolate–imidizolium 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

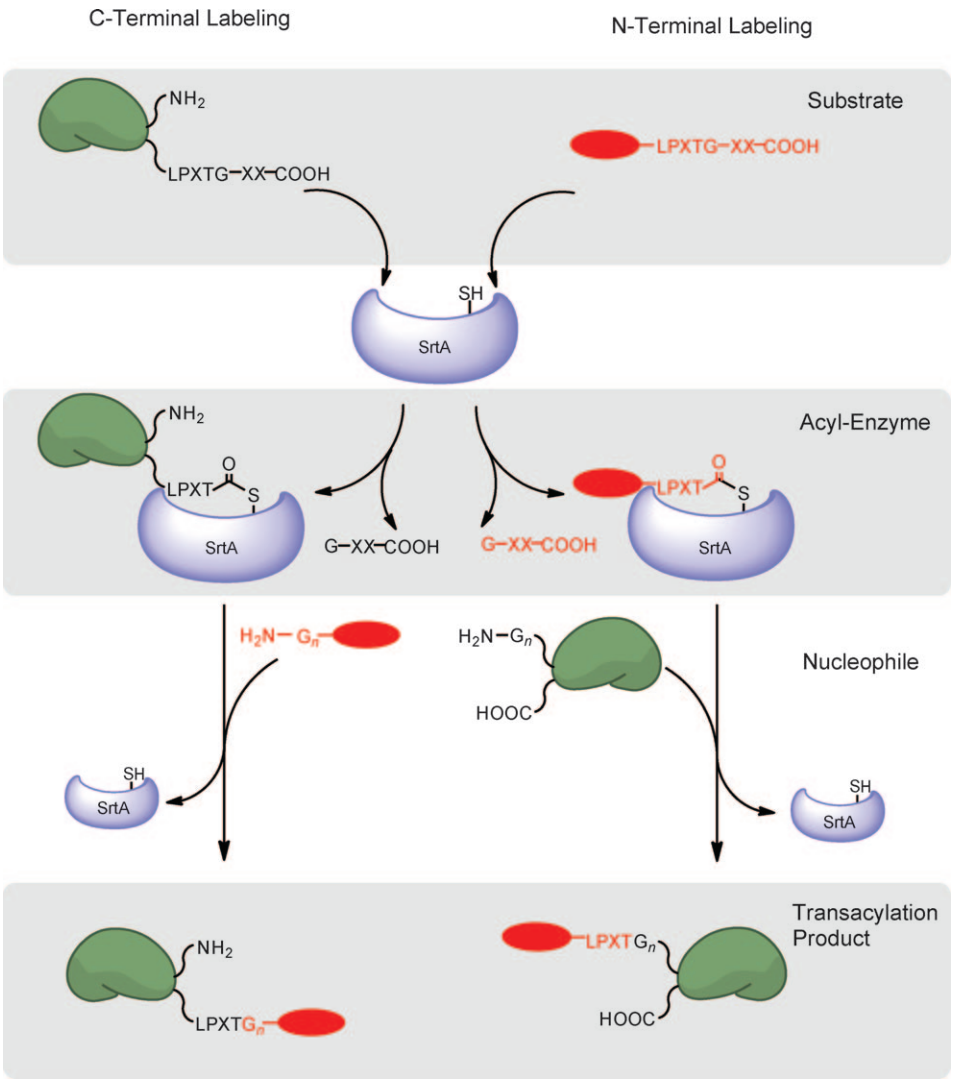


Figure 1. Site-specific C- and N-terminal labeling scheme using sortase A. C-Terminal labeling (left) and N-terminal labeling (right) proceed through a substrate-recognition step (top), followed by generation of a thioacyl intermediate (middle) and resolution of the acylated enzyme by an exogenously added nucleophile (bottom). See text for details.

catalytic cysteine residue.[43] 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.[44,45] The cysteine residue attacks the amide bond between the threonine and glycine residues in the sorting motif. The protonated imidizolium 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.

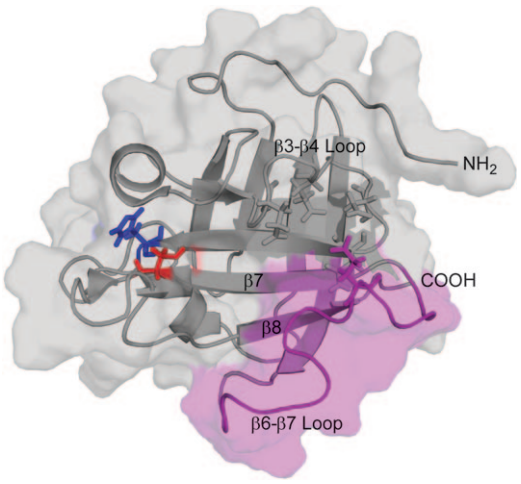


Figure 2. Structure of sortase A from *Staphylococcus auerus* deduced by NMR spectroscopy (pdb code: 1IJA).[36] The active-site cysteine residue (Cys184) is in red and the active-site histidine residue (His120) is in blue. The β7 and β8 strands that form the floor of the active site are labeled and the β6-β7 loop involved in substrate recognition[93] is in purple. Residues that coordinate calcium are shown as sticks.[42]

acts as a general acid for the departing αNH_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.[46]

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 though the normal cell-wall sorting pathway. In this manner, the enzyme alkaline phosphatase has been anchored to the cell wall of *Staphylococcus aureus*,[47] the E7 protein of Human papilloma virus 16 (HPV16) has been displayed on *Streptococcus gordonii*, a commensal microbe in the oral cavity,[48] and α-amalyase has been affixed to the peptidoglycan of *Bacillus subtilis*, helped by coexpression of the sortase gene from *Listeria monocytogenes*.[49] 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.[50] 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.

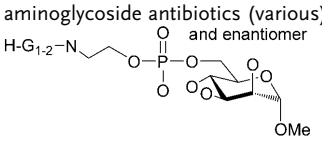
3. C-Terminal Labeling

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).[34] 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.[51]

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.[52] 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,[53] and large polyethylene glycol chains[54] 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.[55] 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[a]	Labeling site	Property endowed	Reference(s)
1	H-G_5K(biotin)L-OH	C terminal	biophysical handle	[32, 55]
2	H-G_5K(ANP)K(biotin)L-OH	C terminal	biophysical handle/photocleavage	[32]
3	H-G_5K(phenylazide)K(biotin)G-OH	C terminal	biophysical handle/photo-crosslinker	[32]
4	H-G_3K(FITC)-NH_2	C terminal	fluorescence	[32]
5	H-G_3K(K(TAMRA))-NH_2	C terminal	fluorescence	[32, 71]
6	H-G_3YC(biotin)-NH_2	C terminal	biophysical handle	[33]
7	H-G_3YC(Alexa 488)-NH_2	C terminal	fluorescence	[33]
8	H-AA-Ahx-K(K(TAMRA))-NH_2	C terminal (*S. pyogenes*)	fluorescence	[56]
9	H-G_3K(C12-C24)-NH_2	C terminal	lipidation	[64]
10	H-G_3K(1-ad)-NH_2	C terminal	hydrophobicity	[64]
11	H-G_3WK(cholesterol)-NH_2	C terminal	lipidation	[64]
12	D-Tat (1st residue is G)	C terminal	cell penetration	[53]
13	H-G_2Y-PTD_5-NH_2	C terminal	cell penetration	[53]
14	(H_2NRRQRRTSKLMKRAhx)$_2$KYK(GG-NH_2)-NH_2	C terminal	cell penetration	[53]
15	H-G_3K(folate)-NH_2	C terminal	folic acid	[53]
16	H_2N-PEG	C terminal	inert polymer	[54]
17	H-G_3K(PEG)-OH	C terminal	inert polymer	[54]
18	H-G_3-MAP-NH_2	C terminal	cell penetration	[52]
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_2	N terminal	biophysical handle	[57]
23	Alexa 647-LPETGG-NH_2	N terminal	fluorescence	[57]
24	Alexa 488-LPETGG-NH_2	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_2	N terminal	fluorescence	[50]
28	biotin-GGLPETG-NH_2	N terminal	biophysical handle	[50]
29	N_3-ALPETG-NH_2	N terminal	handle for bioorthognal 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,[32] and osteoclast differentiation factor (ODF)[33] 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[56] as well as the complete LPXTG sortase recognition motif can be used as scaffolds for such probes.[57] 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 αNH_2 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.[57] This system was later used to install reporter fluorphores 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. PFAR receptors with key mutations were then shown to traffic aberrantly.[58] 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-

Table 2: Examples of proteins labeled by sortase A transpeptidation.

Substrate	Solution/cell surface	Labeling site	Label(s)[a]	Reference(s)
H-2K[b]	solution	C terminal	1,2,3,4,5	[32]
CD154	cell surface	C terminal	1,5	[32]
neuramnidase	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]
$(His)_6$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]
$LPETG_5$-ECFP-TM	cell surface	N terminal	22,23	[57]
$LPETG_5$-PAFR	cell surface	N terminal	24	[58]
G_3-/G_5-CTXB	solution	N terminal	25,26	[56]
G_3-eGFP	solution	N terminal	26	[56]
G-UCHL3	solution	N terminal	26	[56]
S. aureus surface peptidoglycan	cell surface	N terminal	27,28,29	[50]
eGFP	solution	N and C terminal	26 and8	[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[59] ($SrtA_{Strep}$) recognizes and cleaves the LPXTA motif and accepts alanine-based nucleophiles. It also cleaves the $SrtA_{Staph}$ (Sortase A from *Staphylococcus aureus*) LPXTG motif, albeit with reduced efficiency. In contrast, the $SrtA_{Staph}$ 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 $SrtA_{Staph}$ labeling were exposed after proteolytic cleavage by thrombin. This step avoids protein oligomerization, likely to occur during the C-terminal labeling step with $SrtA_{Strep}$.[56]

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.[60] 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.[61] Various peptides (CD52 fragment, Mucin 1 peptides) and small proteins (CD24) have been attached to GPI mimics with trisaccharide cores.[62,63] 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.[64] A glycine-based scaffold was modified with a panel of linear alkyl chains (C_{12}–C_{24}) 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_{22} 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_3-GB1) onto the C terminus of the Vav SRC homology 3 domain (SH3), a domain that is nearly insoluble by itself at pH 7.[65] The structure of the attached $^{13}C/^{15}N$-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.[66] 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.[67] 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 stereochemcially homogeneous preparation.[68,69] Sortase has been applied to the construction of G-protein-coupled receptor (GPCR) mimics through a combined recombinant, enzymatic, and chemical synthesis (CRECS) strategy.[70] 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.[54,71] 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.[72–76]

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.[54] The attachment of adhesion proteins from Gram-positive bacteria to fluorescent glycine-derivatized polystyrene beads was done in a similar manner.[77] 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.[78] The directional anchoring of proteins onto triglycine-modified caroboxymethylated dextran-based Biacore sensor chips for use in surface plasmon resonance has also been accomplished.[79] 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 β-1,4-galactosyltransferase (rhGalT) and *Helicobacter pylori* α-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.[80]

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^{2+} and triglycine to yield highly pure protein with one additional glycine residue[81] (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 strepavidin resin in the presence of calcium chelators to prevent premature cleavage.[82] 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 ($>40\,mg\,L^{-1}$ 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.[83]

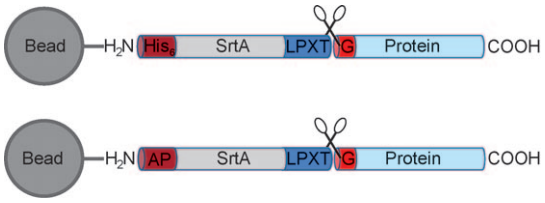


Figure 3. Protein purification using sortase A. Recombinant expressed proteins are produced as fusion proteins containing either a hexahistidine tag (top)[81] or a biotin acceptor peptide (bottom)[82] followed by the catalytic core of sortase, the LPXTG tag, and the protein of interest. Addition of Ca^{2+} 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.[84] The pilin-building sortases, for example, accept the ε-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,[85,86] while SrtB from *Bacillus anthra*cis cleaves the NPKTG sequence[87] and accepts *meso*-diaminopimelic acid as a nucleophile.[88–90] 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:[91] 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 β6-β7 loop,[92–94] 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,[88,89] 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.[95] 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,[95] and yields circular proteins intracellularly in yeast, mammalian cells, and prokaryotes. The limiting step for simple C-terminal labeling is thus delivery of an exogeneous 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 α-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 aminooxy 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 $SrtA_{Staph}$ and $SrtA_{Strep}$ 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: April 27, 2011

[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 address: 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. 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. 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**, Chap. 15, Unit 15.3.
[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. Kazmiercak, 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. 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. 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. Kurogochi, 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. Rothemund, K. Fechner, M. Bienert, M. Beyermann, *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.
[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.

