**Protein Modification**

Efficient N-Terminal Labeling of Proteins by Use of Sortase**

Daniel J. Williamson, Martin A. Fascione, Michael E. Webb, and W. Bruce Turnbull**

Protein labeling is a pivotal technique in molecular and cell biology. Strategies include derivatization of cysteine residues,[1] labeling lysine or N-terminal amino groups with activated esters, periodate or PLP-mediated oxidation of the N-terminus for oxime ligation,[2] and native-chemical ligation.[3] Each of these methods has its own associated challenges: Selective labeling of a single cysteine residue frequently requires rounds of site-directed mutagenesis to introduce the labeling site and/or remove other cysteine residues, and selective labeling of the N-terminal amino group requires careful control of pH to ensure lysine residues are not also modified.[4] Other modern methods for chemoselective labeling often require the introduction of specific recognition sequences[5] or nonnatural amino acids into the protein to be labeled.[6] In most cases, a substantial excess of the labeling reagent is necessary to ensure complete conversion to the product. We report a method for chemoselective N-terminal labeling of recombinant proteins in quantitative yield using depsipeptide substrates for the transpeptidase sortase A. The method does not require engineering of the protein sequence beyond that typically used in contemporary recombinant protein purification strategies. Unlike previous approaches,[7] the method requires only a single N-terminal glycine residue in a sterically unhindered position, a minimal excess of the labeling reagent and substoichiometric quantities of transpeptidase.

Sortase A (SrtA) catalyzes the reversible attachment of virulence factors to the cell walls of Gram positive bacteria by C-terminal modification of proteins at an LPXTG recognition sequence.[8] The enzyme catalyzes the covalent attachment of the LPXT motif to a cysteine residue in the catalytic site to form a thioester intermediate. An N-terminal oligoglycine motif in the peptidoglycan can then react with this intermediate to covalently attach the substrate to the cell wall. SrtA has been exploited extensively for the C-terminal modification of proteins.[7b, 9] However, this method has

certain constraints: the LPXTG sequence must be engineered into the protein and excess nucleophilic labeling reagent is required to push the equilibrium toward formation of product as the transpeptidase reaction is reversible (Scheme 1a).
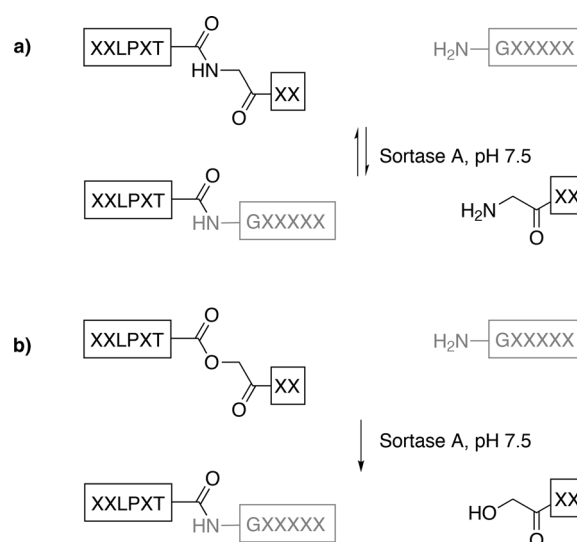
Scheme 1. Sortase A ligation employing a) peptide and b) depsipeptide substrates.

N-terminal labeling of proteins using SrtA has the potential advantage that only a single N-terminal glycine is required on the protein.[10] Since many commercial expression plasmids incorporate protease recognition sequences which yield an N-terminal glycine after cleavage, such an approach is potentially widely applicable. However, reversibility of the SrtA-catalyzed reaction is a greater problem for N-terminal labeling of proteins.

We chose to investigate whether the reaction could be made irreversible by using depsipeptide substrates (Scheme 1b). We anticipated that the hydroxyacetyl byproduct would not be a substrate for the reverse reaction, thus rendering the labeling reaction irreversible. A similar strategy has been applied previously to subtiligase,[11] but overall yields were dependent on specific recognition sequences in both peptides and an excess of depsipeptide was necessary to drive the reaction to completion. Ploegh and co-workers have reported the use of a methyl ester substrate with SrtA;[7a] however, their method employed stoichiometric quantities of SrtA and excess methyl ester.

We first chose to establish an assay to evaluate the efficiency of the SrtA reaction with various peptide substrates. Three classes of labeling substrate **1**, **2** and **3** were synthesized (Figure 1a); these peptides share a common YALPET sequence followed by a single glycine or glycine-

[*] D. J. Williamson, Dr. M. A. Fascione, Dr. M. E. Webb,
Dr. W. B. Turnbull
School of Chemistry and Astbury Centre for Structural Molecular
Biology, University of Leeds, Leeds LS2 9JT (UK)
E-mail: m.e.webb@leeds.ac.uk
        w.b.turnbull@leeds.ac.uk
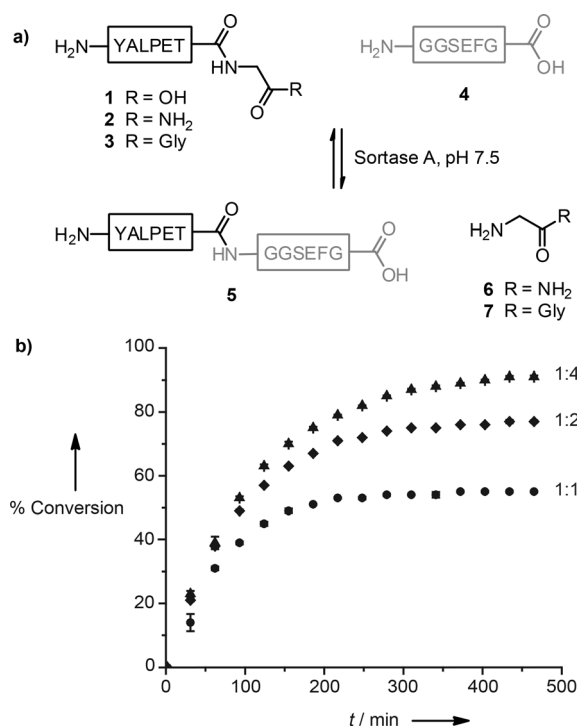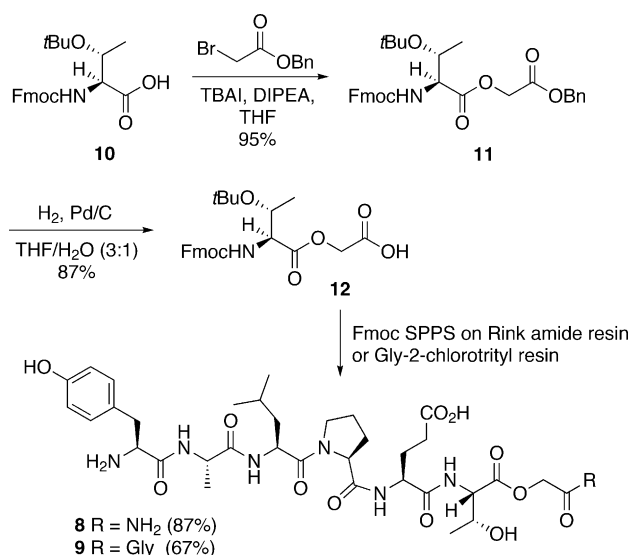Homepage: http://www.chem.leeds.ac.uk/MEW/

Figure 1. a) Model reactions to investigate ligation efficiency for a variety of peptide substrates. Peptide **1** was not turned over by SrtA. b) Time-course of SrtA-mediated ligation of **3** and **4** to generate **5**. Increasing the ratio of the **3** to **4** is sufficient to drive the reaction to completion. ●, ◆, and ▲ denote the ratio of **3** to **4** as indicated. The reaction was monitored by HPLC using 11 μM SrtA to catalyze the reaction of 500 μM **3** with the corresponding ratio of the nucleophilic acyl acceptor **4** in 50 mM HEPES, 5 mM CaCl$_2$ 150 mM NaCl pH 7.5.

Scheme 2. Synthetic strategy for preparation of depsipeptides for use in SrtA-mediated ligation.

SrtA-mediated ligation of each depsipeptide with GGSEFG peptide **4** was followed by using HPLC (Figure 2). When one equivalent of either depsipeptide **8** or
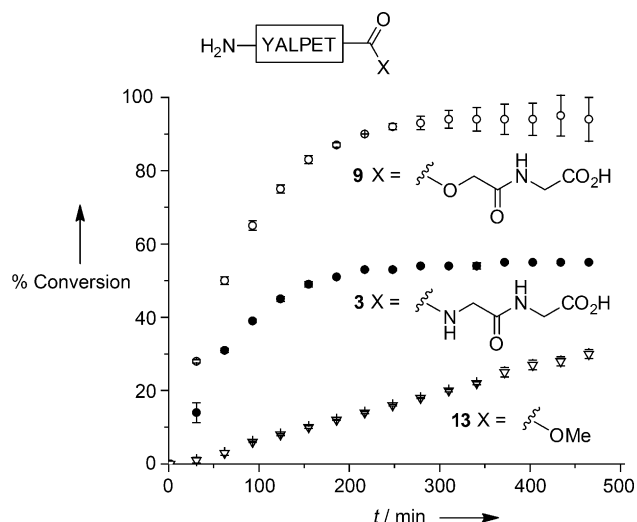

Figure 2. Time-course of SrtA-mediated ligation of ligation of **3**, **9** and **13** with **4** at a 1:1 ratio at 37°C. 500 μM **4** and 500 μM **3**, **9** or **13** with 11 μM SrtA in 50 mM HEPES, 5 mM CaCl$_2$ 150 mM NaCl pH 7.5.

amide, or two glycines, respectively. All three peptides were tested as substrates for ligation to a short peptide with the sequence GGSEFG **4** using SrtA in aqueous buffer (Figure 1a). HPLC analysis using an authentic sample of product **5** as a standard showed that only peptides **2** and **3** act as substrates for SrtA, as indicated previously.[12] Neither reaction achieved complete conversion to product **5** as this peptide is itself a substrate for SrtA and can thus react with glycine amide **6** or diglycine **7** to reform the starting peptides. The reaction mixture therefore comes to equilibrium after 50% conversion (Figure 1b). Increasing the ratio of the "nucleophilic" peptide **4** to the "electrophilic" acyl donor peptides **2** and **3** leads to an increase in conversion but the system still goes to equilibrium (Figure 1b and Supporting Information).

Depsipeptides **8** and **9** were synthesized by Fmoc solid phase peptide synthesis (SPPS) using Fmoc-protected TG depsipeptide **12** (Scheme 2). First, alkylation of Fmoc *tert*-butyl-protected threonine **10** with benzyl bromoacetate in the presence of tetrabutylammonium iodide yielded the triprotected depsipeptide **11**. Hydrogenolysis of the benzyl group gave carboxylic acid **12** which was activated using HCTU for attachment to NovaGel Rink amide resin or glycine-loaded 2-chlorotrityl resin. Standard Fmoc-SPPS protocols were used to extend these depsipeptides to provide compounds **8** and **9** in 87% and 67% yield, respectively.

9 was used, **4** was almost quantitatively transformed to the ligation product **5** (Figure 2 and Supporting Information). The presence of 2 mol% SrtA in the reaction mixture prevents the reaction going to completion as the equilibrium between the product and thioester intermediate remains significant. In contrast, the corresponding methyl ester **13** reacted slowly under these conditions, leading to only 30% product formation after 8 h. When the amount of depsipeptide **9** was increased to 1.5 equiv, transformation of **4** to **5** became rapid and quantitative (see Supporting Information).

As depsipeptides proved effective for labeling model peptides, we sought to apply the method to labeling proteins. SrtA-mediated ligations are typically carried out using peptides containing an N-terminal oligoglycine motif; however, this is not frequently observed in proteins. A single N-terminal glycine is, however, commonly produced in recombinant proteins after cleavage with proteases such as thrombin or tobacco etch virus (TEV) protease. Depsipeptide **14** containing a dansyl lysine residue was synthesized using the same methodology as for **8** and **9**. Initial labeling experiments were performed using a variant of the human mannose binding protein (ManBP).[13] This trimeric protein has a single N-terminal glycine at the end of a three-stranded α-helical bundle. The substrate protein was labeled quantitatively within 4 h when incubated with 1.5 equiv of the depsipeptide labeling reagent per protomer (Figure 3). The degree of labeling was confirmed by electrospray mass spectrometry (ESMS) of the reaction mixture, which indicated complete conversion to the labeled protein (Figure 3c). A small quantity of acylated SrtA can also be observed in the SDS-PAGE gels of the crude reaction mixtures, corresponding to 10 mol% SrtA present in the reaction mixture. SrtA can be easily removed by affinity purification to yield a pure reaction product. ManBP was also labeled successfully with a fluorescein-modified depsipeptide (see Supporting Information). When working at low protein concentrations, prolonged incubation with SrtA can lead to hydrolysis of the label;[10] however, if necessary, increasing the quantity of label to 2–3 equiv can still allow complete conversion to the product.

To confirm the generality of our depsipeptide method we also labeled a sample of the mouse pumilio-2 Puf RNA-binding domain[14] in which an N-terminal GT sequence had been produced by TEV protease cleavage. The protein was quantitatively labeled using 1.5 equivalents of labeling reagent **14** demonstrating the generality of this approach. We also investigated the labeling of both myoglobin and the fly pumilio RNA-binding domain. In both cases protein labeling was unsuccessful; however, short peptides corresponding to the N-terminal sequences of myoglobin, mouse pumilio and fly pumilio could be successfully modified (see Supporting Information). All three peptides displayed very similar reaction kinetics that were only slightly slower than those displayed for diglycyl peptide **4**. We attribute the differences in protein reactivity to variation in steric bulk in the vicinity of the labeled glycine. In the case of fly pumilio, the glycine is only one residue removed from the globular domain of the protein with the sequence GS (c.f. GTG for the mouse paralog). This suggests that a minimum length of flexible peptide is required in order to ensure that labeling can occur.

The depsipeptide substrates allow rapid labeling of both peptides and proteins. In conventional SrtA-mediated ligations, the rate-determining step is the initial attack of the enzyme to form an acyl–enzyme intermediate;[10] this is then attacked by the nucleophile to form a product that is itself a substrate for SrtA. The efficiency of any given substrate is therefore controlled by the rate of its turnover relative to that for the product. The Michaelis constants for peptide acyl donors are typically in the same range as the substrate
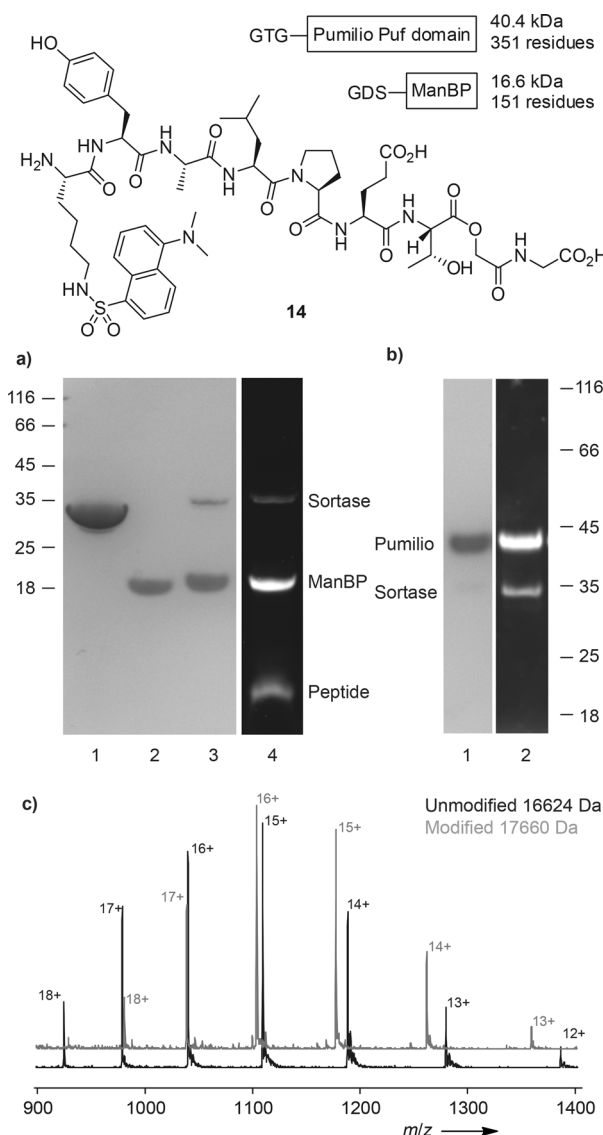

Figure 3. Use of **14** in covalent modification of proteins. a) Labeling of mannose-binding protein with **14**. Incubation of ManBP with 0.1 equiv SrtA and 1.5 equivalents of **14** led to quantitative labeling in 4 h: lane 1, SrtA; lane 2, 60 μM ManBP; lane 3, 60 μM ManBP + 90 μM **14** + 6 μM SrtA; lane 4, fluorescence image of lane 3. b) The same protocol led to quantitative labeling of mouse pumilio-2 Puf domain: lane 1, 20 μM pumilio + 30 μM **14** + 2 μM SrtA; lane 2, fluorescence image of lane 1. c) MS analysis of ManBP before (black) and after (gray) modification demonstrates quantitative labeling of protein.

concentrations used in this study;[10,15] the relative rate of reaction is therefore determined by the specificity constant, k_{cat}/K_m. We suggest that this factor accounts for the success of the depsipeptides under the conditions used in this study. The rate of acylation by an ester substrate will be greater than that for an amide, thus increasing k_{cat} for both the depsipeptide and the methyl ester. However, while the affinity of the depsipeptide should be similar to that of a peptide substrate, the methyl ester will presumably bind less tightly. Overall the relative rate of reaction (as determined by k_{cat}/K_m) for the methyl ester should be less than that for the product, and product inhibition therefore leads to the observed low overall

rate of reaction. The rate of reaction for the depsipeptide will be higher than that for the product and rapid conversion is thus observed.

In conclusion, we have described the synthesis and application of depsipeptide substrates for sortase-mediated ligation at the N-terminus of proteins. These substrates enable labeling of peptides and proteins with virtually equimolar quantities of each coupling partner and with substoichiometric quantities of sortase, as long as the N-terminal glycine residue is sterically accessible. This approach is ideally suited to the use of high-value reagents and minimizes the requirement for post-reaction purification. While the method introduces an LPET sequence at the N-terminus of the labeled protein, we would not anticipate that this modification should alter the biochemical behavior of a protein any more than introduction of a fluorescent tag by any other method. We thus expect that it will be of widespread utility for the N-terminal modification of proteins.

Experimental Section

Full experimental details are in the Supporting Information.

Protocol for protein labeling. All stock solutions were prepared in 50 mM HEPES, 5 mM CaCl$_2$, 150 mM NaCl pH 7.5 at the following concentrations: acyl donor **14** 2 mM, ManBP 120 µM, Sortase A 304 µM. In a 100 µL assay at a ratio of 1.5:1, ManBP (50 µL, 60 µM), **14** (4.5 µL, 90 µM) and Sortase A (2 µL, 6 µM) were added to HEPES buffer (43.5 µL) and mixed well. The solution was incubated at 37 °C. Reaction progress was monitored by SDS-PAGE and ESMS analysis.

[1] a) J. M. Chalker, B. G. Davis, *Curr. Opin. Chem. Biol.* **2010**, *14*, 781–789; b) D. P. Gamblin, P. Garnier, S. van Kasteren, N. J. Oldham, A. J. Fairbanks, B. G. Davis, *Angew. Chem.* **2004**, *116*, 846–851; *Angew. Chem. Int. Ed.* **2004**, *43*, 828–833.

[2] a) J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi, M. B. Francis, *Angew. Chem.* **2006**, *118*, 5433–5437; *Angew. Chem. Int. Ed.* **2006**, *45*, 5307–5311; b) J. Chen, W. Zeng, R. Offord, K. Rose, *Bioconjugate Chem.* **2003**, *14*, 614–618.

[3] a) P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, *Science* **1994**, *266*, 776–779; b) C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem.* **2008**, *120*, 10182–10228; *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074.

[4] D. P. Baker, E. Y. Lin, K. Lin, M. Pellegrini, R. C. Petter, L. L. Chen, R. M. Arduini, M. Brickelmaier, D. Wen, D. M. Hess, L. Chen, D. Grant, A. Whitty, A. Gill, D. J. Lindner, R. B. Pepinsky, *Bioconjugate Chem.* **2005**, *17*, 179–188.

[5] M. J. Hinner, K. Johnsson, *Curr. Opin. Biotechnol.* **2010**, *21*, 766–776.

[6] a) A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609; b) J. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775–782; c) A. J. de Graaf, M. Kooijman, W. E. Hennink, E. Mastrobattista, *Bioconjugate Chem.* **2009**, *20*, 1281–1295.

[7] a) J. M. Antos, G.-L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W.-L. Popp, H. L. Ploegh, *J. Am. Chem. Soc.* **2009**, *131*, 10800–10801; b) T. Yamamoto, T. Nagamune, *Chem. Commun.* **2009**, 1022–1024; c) N. Hirota, D. Yasuda, T. Hashidate, T. Yamamoto, S. Yamaguchi, T. Nagamune, T. Nagase, T. Shimizu, M. Nakamura, *J. Biol. Chem.* **2010**, *285*, 5931–5940.

[8] a) S. K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, *Science* **1999**, *285*, 760–763; b) H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12424–12429.

[9] a) H. Y. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670–2671; b) L. Chan, H. F. Cross, J. K. She, G. Cavalli, H. F. P. Martins, C. Neylon, *PLoS One* **2007**, *2*, e1164; c) M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, *Nat. Chem. Biol.* **2007**, *3*, 707–708; d) S. Pritz, Y. Wolf, O. Kraetke, J. Klose, M. Bienert, M. Beyermann, *J. Org. Chem.* **2007**, *72*, 3909–3912; e) J. M. Antos, G. M. Miller, G. M. Grotenbreg, H. L. Ploegh, *J. Am. Chem. Soc.* **2008**, *130*, 16338–16343; f) F. Clow, J. Fraser, T. Proft, *Biotechnol. Lett.* **2008**, *30*, 1603–1607; g) T. Proft, *Biotechnol. Lett.* **2010**, *32*, 1–10; h) M. W.-L. Popp, H. L. Ploegh, *Angew. Chem.* **2011**, *123*, 5128–5137; *Angew. Chem. Int. Ed.* **2011**, *50*, 5024–5032; i) X. Guo, Z. Wu, Z. Guo, *Bioconjugate Chem.* **2012**, *23*, 650–655.

[10] X. Huang, A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei, G. Ellestad, *Biochemistry* **2003**, *42*, 11307–11315.

[11] a) T. K. Chang, D. Y. Jackson, J. P. Burnier, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12544–12548; b) H. A. I. Yoshihara, S. Mahrus, J. A. Wells, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6000–6003; c) D. J. Suich, M. D. Ballinger, J. A. Wells, W. F. DeGrado, *Tetrahedron Lett.* **1996**, *37*, 6653–6656.

[12] J. W. Nelson, A. G. Chamessian, P. J. McEnaney, R. P. Murelli, B. I. Kazmiercak, D. A. Spiegel, *ACS Chem. Biol.* **2010**, *5*, 1147–1155.

[13] a) S. Sheriff, C. Y. Y. Chang, R. A. B. Ezekowitz, *Nat. Struct. Biol.* **1994**, *1*, 789–794; b) W. I. Weis, M. E. Taylor, K. Drickamer, *Immunol. Rev.* **1998**, *163*, 19–34.

[14] H. T. Jenkins, R. Baker-Wilding, T. A. Edwards, *J. Struct. Biol.* **2009**, *167*, 271–276.

[15] M. L. Bentley, E. C. Lamb, D. G. McCafferty, *J. Biol. Chem.* **2008**, *283*, 14762–14771.