Zitierweise: *Angew. Chem. Int. Ed.* **2024**, *63*, e202310862
doi.org/10.1002/anie.202310862

Protein Modification

Quantitative N- or C-Terminal Labelling of Proteins with Unactivated Peptides by Use of Sortases and a D-Aminopeptidase

Zoe L. P. Arnott, Holly E. Morgan, Kristian Hollingsworth, Charlotte M. E. Stevenson, Lawrence J. Collins, Alexandra Tamasanu, Darren C. Machin, Jonathan P. Dolan, Tomasz P. Kamiński, Gemma C. Wildsmith, Daniel J. Williamson, Isabelle B. Pickles, Stuart L. Warriner, W. Bruce Turnbull, and Michael E. Webb**

Abstract: Quantitative and selective labelling of proteins is widely used in both academic and industrial laboratories, and catalytic labelling of proteins using transpeptidases, such as sortases, has proved to be a popular strategy for such selective modification. A major challenge for this class of enzymes is that the majority of procedures require an excess of the labelling reagent or, alternatively, activated substrates rather than simple commercially sourced peptides. We report the use of a coupled enzyme strategy which enables quantitative N- and C-terminal labelling of proteins using unactivated labelling peptides. The use of an aminopeptidase in conjunction with a transpeptidase allows sequence-specific degradation of the peptide by-product, shifting the equilibrium to favor product formation, which greatly enhances the reaction efficiency. Subsequent optimisation of the reaction allows N-terminal labelling of proteins using essentially equimolar ratios of peptide label to protein and C-terminal labelling with only a small excess. Minimizing the amount of substrate required for quantitative labelling has the potential to improve industrial processes and facilitate the use of transpeptidation as a method for protein labelling.

Introduction

Site-specific labelling of proteins is a widely used tool in both industrial and academic research.[1] Transpeptidases, such as sortase, connectase, subtiligase and asparaginyl endopeptidases, provide an effective strategy for labelling the N- and C-termini of proteins.[2,3] This approach has now been developed and applied to increasing complex systems including cell-surface labelling,[4] in vivo labelling[5] and the production of therapeutic antibody-drug conjugates.[6] Numerous refinements to these methods have been reported including the evolution of enzymes with enhanced[5,7] and altered[8] reactivity, the use of multiple enzymes to enable orthogonal labelling reactions[9] and the development of approaches to the labelling of internal residues.[9d,10] In almost all cases however, an excess quantity of one of the reaction partners is required to enable full conversion of unlabelled to labelled protein.

Transpeptidases recognise a defined peptide motif in their substrate protein or peptide.[2] An active-site cysteine residue reacts with this motif to generate a thioacyl intermediate which then reacts with a substrate peptide or protein to yield the product. The key challenge in such labelling reactions is that they are generally fully reversible (Scheme 1A). The substrate recognition sequence for the enzyme (e.g. LPXT/G for sortase) is still present in the product of the reaction, and the by-products produced during formation of the initial thioacyl intermediate are also

[*] Dr. Z. L. P. Arnott, Dr. H. E. Morgan, Dr. K. Hollingsworth, C. M. E. Stevenson, L. J. Collins, A. Tamasanu, Dr. D. C. Machin, Dr. J. P. Dolan, Dr. T. P. Kamiński, Dr. G. C. Wildsmith, Dr. D. J. Williamson, Dr. I. B. Pickles, Dr. S. L. Warriner, Prof. W. B. Turnbull, Prof. M. E. Webb
School of Chemistry and Astbury Centre for Structural Molecular Biology, University of Leeds
Woodhouse Lane, Leeds LS2 9JT (UK)
E-mail: w.b.turnbull@leeds.ac.uk
m.e.webb@leeds.ac.uk

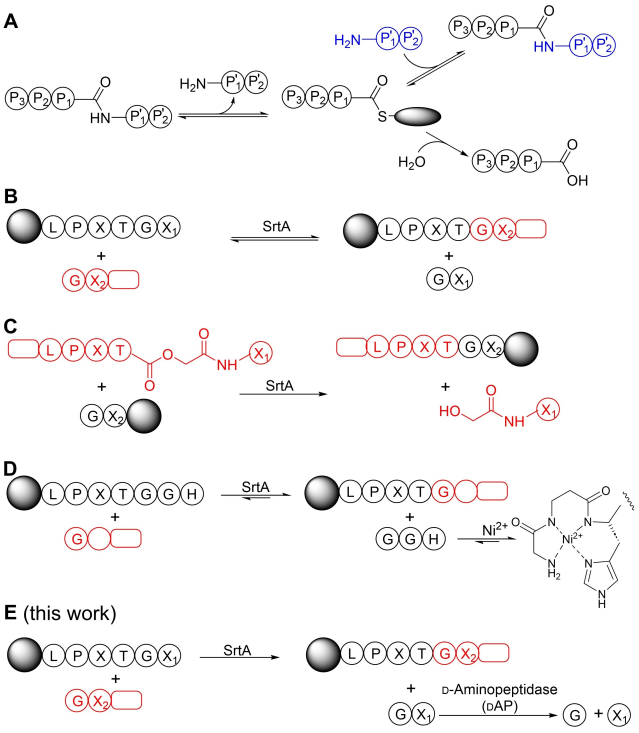
Dr. Z. L. P. Arnott
Present address: Centre for Process Innovation, Central Park, The Nigel Perry Building, 1 Union St, Darlington, DL1 1GL, United Kingdom

Dr. H. E. Morgan
Present Address: Ashfield MedComms, City Tower, Piccadilly Plaza, Manchester M1 4BT, United Kingdom

Dr. J. P. Dolan
Present Address: School of Chemical and Physical Sciences & Centre for Glycoscience Research and Training, Keele University, Keele, Staffordshire ST5 5BG, United Kingdom

Dr. D. J. Williamson
Present Address: Iksuda Therapeutics, The Biosphere, Draymans Way, Newcastle upon Tyne NE4 5BX, United Kingdom

Dr. I. B. Pickles
Present Address: York Structural Biology Laboratory, Department of Biology, University of York, Heslington, York, YO10 5DD United Kingdom



Scheme 1. Peptide and protein labelling reactions catalysed by sortase. A) General reaction scheme for transpeptidation. Sequence-specific recognition. Attack of a catalytic cysteine residue on the peptide bond yields a thioacyl intermediate, which can be cleaved by either aminolysis (to yield a new peptide) or hydrolysis. B) Sortase (SaSrtA) catalyses the transpeptidation of a LPETGX recognition motif in N-terminal glycine-containing peptides and proteins in an equilibrium fashion. C) Strategy for perturbing the equilibria of N-terminal labelling reactions by the use of depsipeptide (ester) substrates. D) Strategy for perturbation of both N- and C-terminal labelling reactions by the use of Ni^{2+}-binding motifs in product peptides. E) Strategy investigated in this study to disrupt these equilibria and yield quantitative labelling. The addition of D-aminopeptidase leads to sequence-specific removal of N-terminal glycines, enabling enhanced labelling using both sortases with broader substrate specificity as well as unmodified peptide. This is controlled by the identity of the amino acid residue following the N-terminal glycine

substrates for the second step. The reversibility of the reaction results in it being under thermodynamic control, which leads to equilibrium mixtures of labelled and unlabelled products (Scheme 1B). Production of pure labelled proteins typically requires a large excess of the labelling peptide and/or removal of unlabelled proteins from reaction mixtures. For N-terminal labelling, we and others have previously reported the use of ester/depsipeptide substrates where the by-product, an alcohol, is no longer a substrate for the enzyme[11] (Scheme 1C). This strategy however is not applicable to C-terminal labelling since it would require incorporation of an ester linkage into an expressed protein. For C-terminal labelling, a variety of alternative strategies have therefore been developed. For example, the physical removal of by-product peptides for C-terminal labelling by the use of centrifugal filtration devices,[12] the use of chemical reagents to react with product peptides[13] or chelation using high concentrations of Ni^{2+} (Scheme 1D).[14] In other work, the substrate preferences of the asparaginyl endopeptidases have been exploited to enable the formation of products that are inefficiently re-processed by the enzyme.[9c] However, at present, generalisable strategies for stoichiometric, quantitative labelling using transpeptidases are limited.

Our previous work using depsipeptides,[9e,11b] methods using immobilised sortases to isolate the thioacyl intermediate[7d] and the mechanical removal of the peptide by-product[12] all suggested that it should be feasible to promote C-terminal labelling if a suitable method to remove the peptide by-product could be found. We hypothesised that a suitable protein modifying enzyme might be capable of selective modification of the peptide N-terminus to facilitate quantitative transpeptidation. Here, we describe the application of a D-aminopeptidase in combination with sortases to enable such quantitative N- and C-terminal labelling of proteins (Scheme 1E).

Results and Discussion

Identification of a sequence-specific aminopeptidase

Our strategy for optimising transpeptidation was initially inspired by the success of our approach for N-terminal labelling using depsipeptide substrates.[9e,11b] In this case, the hydroxyacetyl by-product formed as a result of transpeptidation is not a substrate for the reverse reaction which allows the reaction to proceed to completion. We therefore sought to identify a second enzyme which would selectively degrade a particular peptide by-product to remove it from the reaction equilibrium. Our initial plan was to exploit the substrate promiscuity of sortase A from *Streptococcus pyogenes* (SpSrtA) which is reported to recognise both LPXT/G and LPXT/A as substrate motifs.[9a] We postulated that inclusion of an enzyme which selectively degrades or modifies a peptide with an N-terminal glycine residue would promote labelling of proteins using a substrate peptide with an N-terminal alanine residue. For the second enzyme, we identified D-aminopeptidase (DAP) from *Ochrobactrum anthropi* (*Brucella anthropi*) since this is known to act on both peptides containing N-terminal D-amino acid residues and peptides with N-terminal glycine residues whereas L-amino acid terminated peptides are not substrates.[15] The product of cleavage is a free amino acid and the truncated peptide, and we have previously shown that while glycinamide and peptides terminated with a single glycyl residue are substrates for sortase, glycine is not.[11b]

We obtained recombinant DAP by overexpression as an N-terminal MBP-His-tagged fusion protein and tested activity against a library of peptides terminated with glycine residues. As expected, peptides terminated with glycine residues were substrates for DAP but we observed intriguing sequence selectivity.[9a,16] Such selectivity by DAP has previously been observed for D-Ala-terminated peptide substrates but not fully characterised for all amino acids.[15a] We constructed a library of peptides with the sequence

GXSKYG where the second residue was all naturally occurring amino acids. We initially screened the library (250 μM peptides) at high relative ratios of DAP (10 mol%) for 2 h at 25 °C using LC-MS to identify the formation of product peptides. In this initial low-resolution assay, we observed complete hydrolysis of peptides for a subset of residues (A, D and E, G, M, N, Q, S and T). We later repeated this analysis using lower concentrations of DAP (5 μM with 10 mM peptide) at 37 °C with analysis by HRMS (20 h) and observed the same pattern (Figure 1A) with the exception that we observed some hydrolysis of GC, GH, and GR terminated peptides under these conditions. Broadly those peptides where the second residue contained branched hydrophobic or positively charged side chains were not substrates for DAP; peptides terminated with GV and GP were the poorest substrates and were largely unhydrolysed.

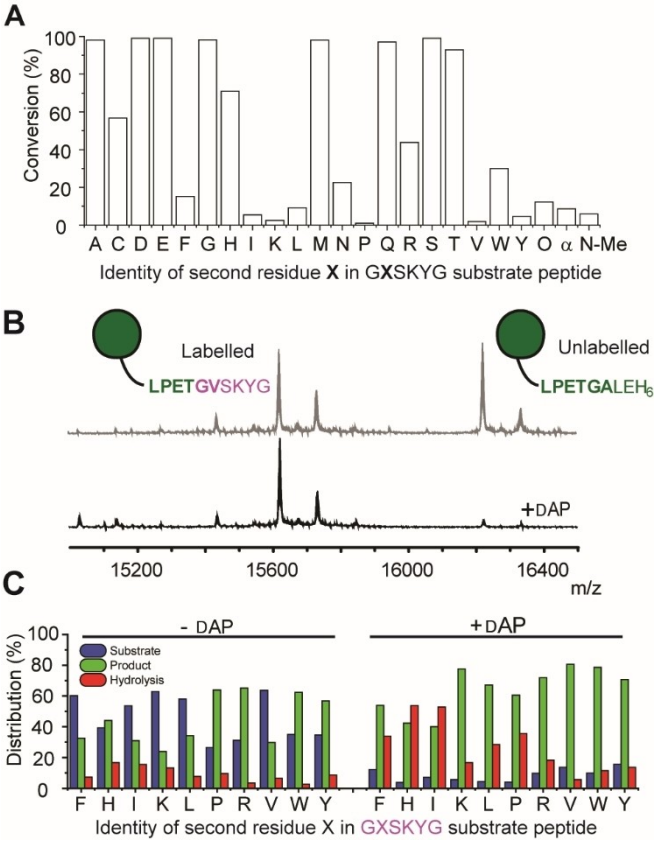


Figure 1. Screening of reaction conditions to enable quantitative labelling. A) Screening substrate specificity of DAP. 10 mM peptide GXSKYG was incubated with 5 μM DAP at 3x°C for 20 h before analysis by HRMS. O: ornithine, α α-methylalanine, N–Me: *N*-methyl alanine. B) Example of reaction MS analysis where the nucleophilic peptide substrate is not a substrate for DAP (e.g. GV-terminated). In the presence of DAP (black), the reaction equilibrium shifts away from the peaks corresponding to unlabeled PanZ towards the labelled species with a small increase in the hydrolysed species. C) Screening of peptide substrates in the presence and absence of DAP enables identification of suitable reaction pairs. 200 μM PanZ-LPETGAH$_6$ was incubated with 400 μM of a model labelling peptide with sequence GXSKYG in the presence of 10 μM Srt7M with and without 20 μM DAP. Green = -labelled, blue = unlabelled, red = hydrolysis.

We also screened a range of non-natural amino acids at the same position (Figure 1A and S1)—in this case we were able to detect hydrolysis of the Gly residue with a variety of D-amino acids (a, f, l, m or v) and linear non-natural amino acids, such as norvaline and norleucine, but observed limited hydrolysis with the lysine homologue ornithine, α-methylalanine (2-aminoisobutyric acid) and *N*-methyl alanine.

Rather than using SpSrtA as originally planned, we hypothesised that we would instead be able to use the selectivity of DAP for the second residue in its substrate peptide to direct ligation of glycine-terminated substrate peptides by sortase. In other words, we could use peptide sequences LPETGX$_1$ and GX$_2$ such that, while the GX$_1$ product peptide is a substrate for DAP, GX$_2$ is not and the equilibrium product will therefore be a peptide with sequence LPETGX$_2$ (Scheme 1E). We initially tested this hypothesis using model peptides and the Ca-independent SaSrtA 7M mutant[5,17] (Srt7M, Figure S2). Using this model system we saw a shift towards product formation in the presence of DAP. Rather than optimising the reaction on peptides, we immediately moved to investigate labelling of a model protein and used an available construct for the PanD regulatory protein, PanZ.[18] This protein was produced with a LPETGA motif inserted prior to a C-terminal His-tag. In the absence of DAP, using 2 equivalents of a GVSKYG co-substrate, we initially observed ≈50% turnover to product after overnight incubation at 25 °C (Figure 1B grey). Addition of DAP increased product formation to 80% (Figure 1B black) with the remaining protein evenly distributed between the unlabelled substrate and a hydrolysis product. In contrast, when a peptide which is a substrate for DAP (e.g. GA-terminated) was used, we only observed either unlabelled protein or the hydrolysis product formed from the thioacyl intermediate (Figure S3C).

We therefore used labelling of PanZ to screen potential peptide substrates by incubating the peptides and PanZ with Srt7M in the presence and absence of DAP (Figures 1C and S4). When the labelling peptide was a substrate for DAP, we only observed the hydrolysed protein with trace unlabelled substrate but in cases where the labelling peptide is *not* a substrate for DAP, we observed a shift in the product distribution from the unlabelled substrate towards the labelled product with variable proportions of the hydrolysis product. From this screen, we selected the GV-terminated peptide together with the LPETGA or LPETGG motif as our preferred substrates for future experimentation.

Optimisation of C-terminal labelling

Our next aim was to optimise the conditions of the labelling reaction. Increasing the pH to above pH 8 (Figure S5) led to >95% conversion to the product, presumably resulting from favoring aminolysis over the hydrolysis reaction at pH above the pKa of the N-terminal amine, consistent with previous results on the pH optimum of sortase labelling.[19] The proportion of the desired product also increased with the number of equivalents of peptide used for labelling, but only marginal enhancement was observed above two equiv-

alents of the labelling peptide with the PanZ substrate (Figure 2B and S6). Variation in the amount of DAP added to the reaction did not affect overall reaction turnover (Figure S7) and while an increase in temperature from 25 °C to 37 °C allowed a reduction in the amount of catalyst, it did not significantly change conversion levels relative to hydrolysis (Figure S8). Finally, we investigated the effect of common co-solvents on the reaction (Figure S9). Addition of DMSO, glycerol and propylene glycol had no effect on the reaction whereas addition of DMA and DMF inhibited, but did not stop, the reaction.

We next investigated modification of the C-terminus of the pentameric cholera toxin B-subunit (CTB) which is widely used as a tool in cell biology and neuroscience.[20] We have previously reported a CTB construct with a C-terminal sortase recognition site, however periplasmic expression yielded heterogeneous substrates in which only 80–90 % of the protomer chains included the desired extension.[21] We therefore generated a new C-terminally His-tagged CTB construct incorporating a LPETGA labelling motif. Cytosolic expression of the unfolded protein into inclusion bodies, followed by dissolution in 8 M urea, affinity purification and subsequent refolding provided homogeneous material for the labelling experiments. In this case, we used automated sampling to monitor the labelling reaction in real time using 2 equivalents of a labelling peptide and observed maximal (>95 %) labelling after approximately 2 h (Figure 2A and S10). Reduction in the number of equivalents of the labelling peptide to only 1.5 equivalents yielded essentially identical results (Figure 2B, 2C). In this case the overall level of labelling was marginally higher than that observed for PanZ, which we attribute to the longer linker between the globular protein domain and the LPETG recognition motif for the CTB protein.

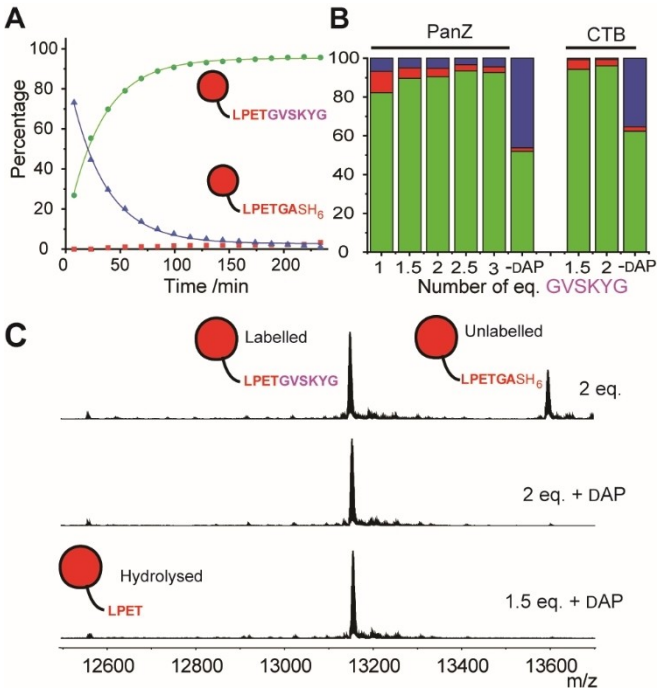


Figure 2. Screening of reaction conditions to enable quantitative labelling. A) Time dependence of C-terminal labelling of the pentameric CTB (100 μM) using 2 equivalents of a labelling peptide, 25 μM Srt7M and 10 μM DAP (green labelled, blue unlabelled, red hydrolysed). B) Peptide concentration dependence of C-terminal labelling of PanZ (200 μM) and CTB (200 μM) after 2 h using 50 μM Srt7M, 20 μM DAP and the indicated ratio of labelling peptide. After 4 h, the proportion of hydrolysis significantly increased (see Figure S6). C) MS analysis of C-terminal labelling of CTB using 1.5 and 2 equivalents of labelling peptide as shown in (B).

Optimisation of N-terminal labelling

Having demonstrated effective C-terminal labelling using this coupling enzyme approach, we investigated its application to N-terminal labelling and compared it to the use of depsipeptide substrates. We tested the enzyme combination using protein constructs in which a GlyVal-labelling motif was added to the N-termini of the expressed proteins. We initially investigated labelling of a pentameric GVG-CTB using mass spectrometry to follow the reaction using a single equivalent (100 μM) of the labelling reagent (Figure 3A and 3B) and 2 mol % Srt7 M in the presence (black) and absence (grey) of 10 mol % DAP. Addition of DAP to the reaction mixture shifted the reaction equilibrium from a statistical 1:1 distribution between labelled and unlabeled products towards only the labelled product. We next compared

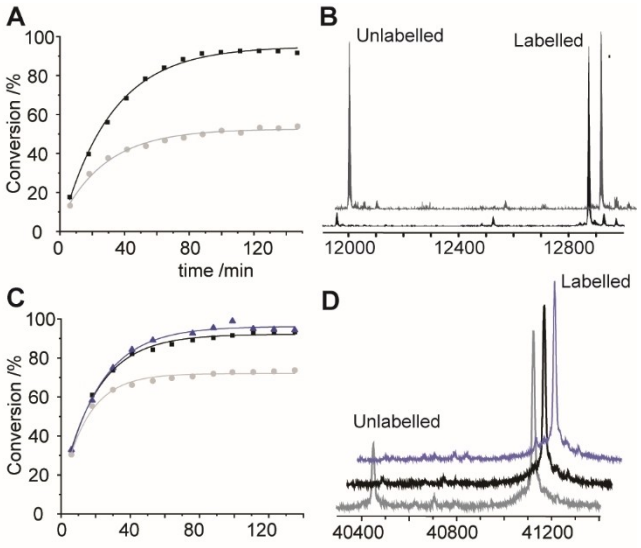


Figure 3. Optimisation of conditions for N-terminal labelling of proteins and comparison with the use of depsipeptide substrates at 25 °C. A ESMS analysis of time-dependent labelling of 100 μM GVG-CTB by Srt7M (2 μM) with 1.2 equivalents of (FITC)-peptide with (black) and without (grey) DAP (10 μM). B) Comparison of N-terminal labelling of GVG-CTB labelling with Srt7M and a fluorescent peptide in the presence (black) and absence (grey) of DAP after 2 h. C) Time-dependent labelling of 100 μM GVG-MBP labelling using Srt7M (5 μM) and 3 equivalents of a peptide substrate in the presence (black) and absence (grey) of DAP (10 μM), or with a depsipeptide substrate (blue). D) MS analysis of MBP labelling using 3 equivalents of labelling reagents after 2 h shows almost complete conversion into labelled product.

labelling using depsipeptides to labelling using the combination of peptides and DAP (Figure S11A) and saw essentially identical behavior using the two approaches. Variation in the concentration of the labelling reagent (Figures S11B and S12) showed that as little as 1.2 equivalents of the peptide label were sufficient to enable ≈95% labelling of this protein at 25°C. Both the depsipeptide and new coupled enzyme approach led to near-quantitative labelling within 2 h. However, the new approach is more robust due to the stability of the peptide substrates in solution in comparison to depsipeptides which are subject to non-enzymatic hydrolysis over time. Labelling of the model maltose-binding protein construct GVG-MBP[9e] revealed a similar pattern (Figure 3C and 3D), with the addition of DAP again enabling near quantitative N-terminal labelling using an unactivated peptide, equivalent to that observed using depsipeptide substrates.

Generation of a chimeric, bifunctional catalyst

Given that addition and removal of two enzymes increases the chance for product contamination and that a single reagent is simpler to handle and monitor, we investigated the combination of both enzymatic activities into a single chimeric polypeptide. Our approaches to N- and C-terminal labelling used variable ratios of the two enzymes but DAP activity never appeared to be a limiting factor—we therefore anticipated that fusing together equimolar quantities of the two enzymes would not significantly affect either class of reaction. We generated fusion proteins consisting of both the Srt7M and DAP domains in tandem with his-tag (SrtH7D) or both his-tag and chitin-binding domain (SrtCH7D) affinity purification tags. SrtH7D and SrtCH7D both expressed at high levels and after purification the chimeric proteins proved stable to short-term storage at 4°C as well as storage at −20 and −80°C.

We initially compared the activity of SrtH7D and SrtCH7D to the separate component proteins for C-terminal labelling of PanZ and observed successful labelling with a slight increase in the level of hydrolysis (Figure S13). We therefore tested C-terminal labelling of CTB using SrtH7D, and in this case observed essentially identical kinetic behaviour to the use of the separate components (Figure 4A). SrtCH7D was also highly effective, and we were able to detect the modified CTB with high levels of labelling using just 2 equivalents of the peptide without further optimisation (Figure 4B). We next used this approach to C-terminally modify MBP, creating a new C-terminal modified MBP-LPETGAH$_6$ construct for that purpose. Once again, we saw essentially quantitative labelling using a small excess of the labelling peptides (Figures 4C, S14 and S15) with only a small difference in labelling efficiency between 1.5 and 3 equivalents of the labelling peptide.

For N-terminal labelling, we initially compared the activity of the SrtH7D/peptide combination with the Srt7M/depsipeptide strategy to label MBP—and saw comparable results using the two approaches and 3 equivalents of a labelling reagent (Figures 4D and S16). We next used GVG-CTB to investigate the concentration dependence of the labelling reaction, comparing the labelling reaction between the combined SrtH7D construct and the separate components at equimolar concentrations (Figure S17). In this case, we observed near quantitative labelling at 25°C using 1.5–2 equivalents of the labelling peptide, which is consistent with the results for the independent components. Finally, we generated a Titin I27 domain (a model protein commonly used in multidomain constructs for mechanical unfolding of proteins) suitable for dual N- and C-terminal labelling. This incorporated a C-terminal sortase motif and a latent N-terminal GVG sequence that could be revealed by action of TEV protease. It was possible to label this protein quantitatively on its C-terminus using 2–3 equivalents of the labelling peptide and as little as 10 mol% SrtH7D over a 2 h time period (Figure 4E and S18). Then, following TEV cleavage to create a new GVG N-terminus, the protein was labelled using an orthogonal LPESoG depsipeptide substrate[9e] and Srt4S-9 (SrtLPXSG) to yield a double labelled protein (Figure 4E).

Further optimisation of N-terminal labelling

Having demonstrated that we can achieve high level N- and C-terminal labelling, we wanted to address three further questions. Can we further enhance the efficiency of labelling, can we use a different sortase in combination with D-aminopeptidase and can we use this method to achieve protein fusion? Protein labelling using the combination of Srt7M and D-aminopeptidase at 25°C is already rapid and nearly equimolar in terms of protein and peptide substrate. For N-terminal labelling, a small excess of peptide is sufficient to enable quantitative labelling however for C-terminal labelling, protein hydrolysis remains a challenge even at relatively high ratios of label to protein. In practice, even though the sortase selectively catalyses aminolysis over hydrolysis, since the latter reaction is irreversible, a small amount of the hydrolysis product is still seen. This problem can be counteracted by carrying out reactions at sufficiently high substrates concentration, however this is not practical for some substrates.

We therefore sought to determine whether we could use temperature to control this side reaction. As we had observed more hydrolysis at 37°C than at 25°C, we decided to investigate the reaction at 4°C, with the hypothesis that the selectivity for aminolysis over hydrolysis would be higher as the temperature decreased. We tested N-terminal labelling of the GVG-CTB construct using 0.02 eq SrtH7D and 1.1 eq peptide with overnight incubation at 4, 16 and 25°C (Figure S19). Although slower, reaction at 4°C led to the same level of labelling as observed for higher temperatures. Following overnight incubation, we observed >99% protein labelling at 4°C, whereas at higher temperatures hydrolysis-mediated unlabelling had occurred. Varying the concentration of peptide revealed that only 1.05 equivalents were required to yield essentially quantitative labelling, and a time course experiment revealed that the level of labelling remained stable under these conditions between 10 and 24 h

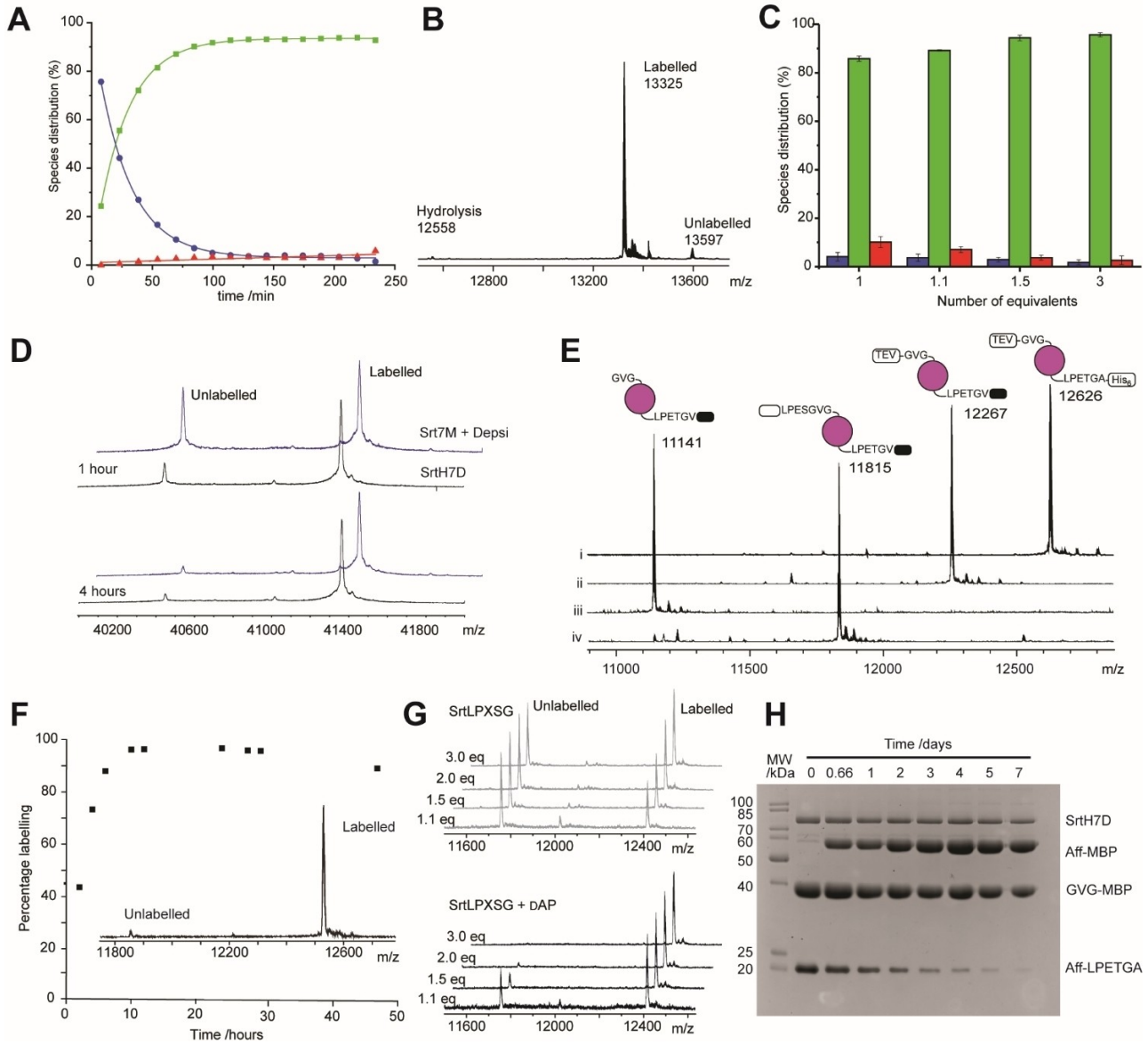


Figure 4. Optimisation of protein labelling using chimeric SrtA-D-aminopeptidase constructs SrtH7D and SrtCH7D at 25 °C and 4 °C and demonstration of enhanced labelling using SrtLPXSG/DAP. A) Time course of C-terminal labelling of 100 μM CTB using 25 μM SrtH7D and 200 μM peptide at 25 °C. B) MS analysis of labelling of 100 μM CTB using 25 μM SrtCH7D after 4 h. C) Peptide substrate concentration dependence of labelling of 100 μM MBP-LPETGA construct after 4 h at 25 °C. (green = labelled, blue = unlabeled, red = hydrolysis). Error bars are standard error $n = 3$. D) Application of SrtH7D *to* N-terminal labelling of GVG-MBP (black) is equivalent to the use of depsipeptide substrates (blue). E) MS analysis of sequential SrtH7D-dependent C-terminal labelling, TEV cleavage and SrtLPXSG/depsipeptide-dependent labelling of I27. i) MS of initial construct; ii) MS analysis of product of C-terminal modification with SrtH7D and peptide GVSKYG; iii) MS analysis of TEV-cleavage: iv) MS analysis of product of N-terminal modification with SrtLPXSG and depsipeptide GABA-YLPESoGG. F) Time course of N-terminal labelling of 100 μM CTB using 105 μM labelling peptide and 2 μM SrtH7D at 4 °C. Inset: MS analysis of labelling at 10 h. G) Labelling of 100 μM GVG-CTB using 5 μM SrtLPXSG with (black) and without (grey) DAP at 4 °C. H) SDS-PAGE analysis of protein fusion between an Affimer (nominal 50 μM) with a C-terminal LPETGA motif and GVG-MBP (50 μM) using SrtH7D (5 μM) at 4 °C.

(Fig 4F) before the peptide label started to be removed by hydrolysis. In practice, these concentrations are within the error of the method used to determine the concentration of reagents suggesting that equimolar N-terminal labelling with sortase has been achieved for the first time. We similarly reinvestigated the C-terminal labelling of both MBP-LPETGA under the same conditions but this time using 0.25 eq of the SrtH7D catalyst. In this case, we saw quantitative labelling of protein with no detectable hydrolysis using 2 equivalents of the labelling peptide (Figure S20) but were not able to reproducibly produce the same success with near equimolar peptide substrates as for the N-terminus.

We have previously demonstrated the orthogonality of the Srt4S-9 (SrtLPXSG) variant to WTSrtA for depsipeptide-mediated labelling[9e] and used those reagents in the

dual labelling approach discussed above. We therefore investigated whether this Ca-dependent variant sortase could be combined with DAP to enable quantitative labelling with peptides. We screened labelling of GVG-CTB at 4 °C in the presence and absence of DAP using a variety of peptide label ratios (Figures 4G and S21). We observed enhanced labelling relative to SrtLPXSG alone in all cases and observed >95 % labelling for both 2 and 3 equivalents of peptide label after 24 h incubation. This suggests that the method should be readily portable to other Ca-dependent and independent variants of sortase with different substrate specificities or to other transpeptidases. Finally, we considered if SrtH7D could be applied to fusion of two globular proteins. We have previously attempted to use sortase variants to enable protein fusion, however at 25 °C we often observed hydrolysis of the N-terminal protein fragment containing the LPETG motif. We therefore investigated fusion at 4 °C using a small excess of the C-terminal fragment. Using these conditions, we observed near complete consumption of an N-terminal fragment (in this case an Affimer protein, Aff-LPETGA with a model C-terminal fragment (GVG-MBP)) over a 7 day incubation period at 4 °C (Figure 4H) using 10 mol % (5 μM) SrtH7D. Slower ligation was observed at 1 and 2.5 μM SrtH7D (Figure S22). Similar trends in successful ligation were observed with dimerisation of MBP (Figure S23) and formation of other fusions from their constituent parts.

Prospects for future applications

The ideal transpeptidation reaction would be one in which two substrates, either peptides or proteins, were mixed in equimolar concentrations in the presence of a catalyst and a quantitative yield of a product is produced. In this work we have demonstrated that combination of a sequence-specific aminopeptidase with a sequence-specific transpeptidase is an effective strategy towards such a reaction. As shown above, the combined broad substrate tolerance of the sortase in terms of the second residue in the nucleophilic substrate and the sequence specificity of the D-aminopeptidase used, enables selective transpeptidation. Given that the two peptidases can selectively form one peptide product preferentially, the key parameter that remains to be optimised is the chemoselectivity of the transpeptidase for aminolysis over hydrolysis.

The established approaches to transpeptidation usually employ a large excess of label to ensure transformation which makes the hydrolysis side-reaction less of a challenge since hydrolysis can be outcompeted by both the substrate and by-product peptides at these concentrations. The critical benefit of the coupled enzyme approach is that it allows us to minimise this excess, however avoiding hydrolysis is then critical to its success. For N-terminal labelling, irreversible hydrolysis of the peptide substrate is of less concern than the hydrolysis of protein is for C-terminal labelling since loss of peptide label to hydrolysis can be readily counteracted using a small excess of the peptide. Using 10 μM protein, it is possible to achieve quantitative N-terminal labelling using only 1.1 or 1.5 equivalent of peptide (Figure S24). However, for C-terminal labelling any hydrolysis leads to an irreversible reduction in overall conversion and an excess of peptide is required to ensure that the peptide/water concentration ratio is maintained. For example, using 10 μM protein, while complete consumption of unlabelled protein is observed at all peptide concentrations, 400 μM peptide is required to fully suppress hydrolysis—this is the concentration of peptide used in labelling higher concentrations of protein (Figures 2C and S25).

At the concentrations used for protein labelling in this study (10–200 μM), we observe approximately 100-fold selectivity for aminolysis over hydrolysis at 25 °C. Given the relative concentrations of labelling peptide (200 μM) and water (55 M) this corresponds to an approximately 2.5×10^6 fold selectivity for aminolysis over hydrolysis at this temperature, which corresponds to an approximately 40 kJ mol^{-1} difference in the energy barrier between the two catalysed pathways. Given this difference, dropping the temperature to 4 °C would be expected to improve the selectivity of reaction 3–4-fold. This estimate is largely consistent with the small enhancement in selectivity we observed. Further improvement in this ratio will require use of a different catalyst. We anticipate that other natural or evolved enzymes could exhibit such selectivity at higher temperatures and that reducing the K_m for the aminolytic substrate will be a key parameter to be optimised to maximise reaction turnover for C-terminal labelling at lower protein substrate concentrations.

In contrast, for N-terminal labelling although a moderate excess of peptide (1.5–2-fold) is required at 25 °C, only a small excess (1.05–1.1-fold) is sufficient to enable essentially quantitative labelling at 4 °C. The small excess of peptide is sufficient to counteract the low level of hydrolysis and the reaction mixtures therefore show stable product distributions for 10–20 h enabling flexibility in protein labelling workflows and potentially allowing downstream applications without removal of the excess label. Finally, we have demonstrated combined application of D-aminopeptidase with two different sortases, however we anticipate that the same strategy could be readily applied to other variant sortases or transpeptidases. For example, we would predict that combination of D-aminopeptidase with OaAEP1 and -NGM, and GL- or GV-terminated substrates would work effectively in a manner complementary to approaches using Ni^{2+}.[14c] In this regard, our approach is very complementary to the work of Xia et al.[22] who have recently reported combination of a glutaminyl cyclase with a peptide asparaginyl ligase, and we anticipate that two approaches could be effectively combined in future applications of orthogonal labelling.

Conclusion

We have demonstrated the combined application of a D-aminopeptidase with sortases enables quantitative labelling of both the N- and C-termini of proteins using a minimal excess of a labelling peptide. The success of labelling is

dependent upon inclusion of a suitable D-aminopeptidase resistant motif (glycine followed by a branched hydrophobic or positively charged) at the N-terminus of the substrate protein for N-terminal labelling or a D-aminopeptidase susceptible motif (glycine followed by principally small, negatively charged, linear and polar residues) after the sortase-recognition motif for C-terminal labelling. At low temperatures, labeling of the N-terminus can be achieved using essentially equimolar quantities of a labelling peptide and at low concentrations of protein substrate. The same conditions can be used to obtain protein fusion between protein subunits. All reagents required are commercially available or readily obtained by bacterial overexpression and purification in high yields. This approach has the potential to widen the scope and application of protein transpeptidation as an approach for protein labelling on a laboratory and production scale.

Supporting Information

Supporting figures and full experimental details may be found in the Supporting Information. The authors have cited additional references within the Supporting Information.[23–25]

Acknowledgements

MEW and WBT acknowledge support from the BBSRC (BB/W01131X/1, BB/R005540/1, BB/P028152/1). ZLPA was funded by BBSRC (BB/R005540/1), KH was funded by EPSRC (EP/R511717/1), HEM, CMES, TPK acknowledge PhD studentship funding from BBSRC (BB/M011151/1), JD acknowledges PhD studentship funding from BBSRC (BB/M011151/1) and Iceni Glycoscience. LJC acknowledges PhD studentship funding from the EPSRC Molecules to Product CDT (EP/S022473/1) AT acknowledges PhD studentship funding from BBSRC (BB/V50953X/1) and GlaxoSmithKline. DCM acknowledges funding from BBSRC (BB/K501049/1, BB/P028152/1 and BB/W01131X/1). GW and DJW acknowledge funding by the BBSRC (BB/M005666/1). IBP was supported by a BBSRC Research Experience Placement. The University of Leeds have filed a patent application covering the material in this manuscript (UK Patent Application Number GB2313834.0).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings will be available in University of Leeds data repository at https://doi.org/10.5518/1441 following an embargo from the date of publication to allow for commercialisation of research findings.

Keywords: Coupling Enzymes · Protein Labelling · Sortase · Transpeptidase

[1] a) J. M. Chalker, G. J. L. Bernardes, Y. A. Lin, B. G. Davis, *Chem. Asian J.* **2009**, *4*, 630–640; b) O. Boutureira, G. J. L. Bernardes, *Chem. Rev.* **2015**, *115*, 2174–2195; c) C. D. Spicer, B. G. Davis, *Nat. Commun.* **2014**, *5*, 4740; d) S. C. Reddington, M. Howarth, *Curr. Opin. Chem. Biol.* **2015**, *29*, 94–99.
[2] H. E. Morgan, W. B. Turnbull, M. E. Webb, *Chem. Soc. Rev.* **2022**, *51*, 4121–4145.
[3] A. C. D. Fuchs, M. Ammelburg, J. Martin, R. A. Schmitz, M. D. Hartmann, A. N. Lupas, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2017871118.
[4] a) T. Yamamoto, T. Nagamune, *Chem. Commun.* **2009**, 1022–1024; b) T. J. Harmand, N. Pishesha, F. B. H. Rehm, W. Ma, W. B. Pinney, Y. J. Xie, H. L. Ploegh, *ACS Chem. Biol.* **2021**, *16*, 1201–1207.
[5] Q. Wu, H. L. Ploegh, M. C. Truttmann, *ACS Chem. Biol.* **2017**, *12*, 664–673.
[6] a) R. R. Beerli, T. Hell, A. S. Merkel, U. Grawunder, *PLoS One* **2015**, *10*, https://doi.org/10.1371/journal.pone.0131177; b) C. M. Le Gall, J. M. S. van der Schoot, I. Ramos-Tomillero, M. P. Khalily, F. J. Van Dalen, Z. Wijfjes, L. Smeding, D. van Dalen, A. Cammarata, K. M. Bonger, C. G. Figdor, F. A. Scheeren, M. Verdoes, *Bioconjugate Chem.* **2021**, *32*, 301–310.
[7] a) I. Chen, B. M. Dorr, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11399–11404; b) L. Chen, J. Cohen, X. Song, A. Zhao, Z. Ye, C. J. Feulner, P. Doonan, W. Somers, L. Lin, P. R. Chen, *Sci. Rep.* **2016**, *6*, 31899; c) H. Hirakawa, S. Ishikawa, T. Nagamune, *Biotechnol. Bioeng.* **2012**, *109*, 2955–2961; d) M. D. Witte, T. Wu, C. P. Guimaraes, C. S. Theile, A. E. Blom, J. R. Ingram, Z. Li, L. Kundrat, S. D. Goldberg, H. L. Ploegh, *Nat. Protoc.* **2015**, *10*, 508–516.
[8] a) M. L. Bentley, H. Gaweska, J. M. Kielec, D. G. McCafferty, *J. Biol. Chem.* **2007**, *282*, 6571–6581; b) B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13343–13348; c) K. Piotukh, B. Geltinger, N. Heinrich, F. Gerth, M. Beyermann, C. Freund, D. Schwarzer, *J. Am. Chem. Soc.* **2011**, *133*, 17536–17539; d) C. J. Podracky, C. An, A. DeSousa, B. M. Dorr, D. M. Walsh, D. R. Liu, *Nat. Chem. Biol.* **2021**, *17*, 317–325; e) L. Schmohl, J. Bierlmeier, F. Gerth, C. Freund, D. Schwarzer, *J. Pept. Sci.* **2017**, *23*, 631–635.
[9] 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. J. Harmand, D. Bousbaine, A. Chan, X. Zhang, D. R. Liu, J. P. Tam, H. L. Ploegh, *Bioconjugate Chem.* **2018**, *29*, 3245–3249; c) F. B. H. Rehm, T. J. Harmand, K. Yap, T. Durek, D. J. Craik, H. L. Ploegh, *J. Am. Chem. Soc.* **2019**, *141*, 17388–17393; d) M. Fottner, M. Weyh, S. Gaussmann, D. Schwarz, M. Sattler, K. Lang, *Nat. Commun.* **2021**, *12*, 6515; e) H. E. Morgan, Z. L. P. Arnott, T. P. Kamiński, W. B. Turnbull, M. E. Webb, *Bioconjugate Chem.* **2022**, *33*, 2341–2347.
[10] a) M. Fottner, A. D. Brunner, V. Bittl, D. Horn-Ghetko, A. Jussupow, V. R. I. Kaila, A. Bremm, K. Lang, *Nat. Chem. Biol.* **2019**, *15*, 276–284; b) S. A. McConnell, B. R. Amer, J. Muroski, J. Fu, C. Chang, R. R. Ogorzalek Loo, J. A. Loo, J. Osipiuk, H. Ton-That, R. T. Clubb, *J. Am. Chem. Soc.* **2018**, *140*, 8420–8423.
[11] a) G. K. T. Nguyen, Y. Cao, W. Wang, C. F. Liu, J. P. Tam, *Angew. Chem. Int. Ed.* **2015**, *54*, 15694–15698; b) D. J. Williamson, M. A. Fascione, M. E. Webb, W. B. Turnbull,

Angew. Chem. Int. Ed. **2012**, *51*, 9377–9380; c) D. J. Williamson, M. E. Webb, W. B. Turnbull, *Nat. Protoc.* **2014**, *9*, 253–262.
[12] L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. Zou, M. Sattler, *J. Biomol. NMR* **2015**, *63*, 1–8.
[13] T. M. S. Tang, D. Cardella, A. J. Lander, X. Li, J. S. Escudero, Y.-H. Tsai, L. Y. P. Luk, *Chem. Sci.* **2020**, *11*, 5881–5888.
[14] a) S. A. Reed, D. A. Brzovic, S. S. Takasaki, K. V. Boyko, J. M. Antos, *Bioconjugate Chem.* **2020**, *31*, 1463–1473; b) D. Row, T. J. Roark, M. C. Philip, L. L. Perkins, J. M. Antos, *Chem. Commun.* **2015**, *51*, 12548–12551; c) F. B. H. Rehm, T. J. Tyler, K. Yap, T. Durek, D. J. Craik, *Angew. Chem. Int. Ed.* **2021**, *60*, 4004–4008.
[15] a) Y. Asano, A. Nakazawa, Y. Kato, K. Kondo, *J. Biol. Chem.* **1989**, *264*, 14233–14239; b) L. Fanuel, C. Goffin, A. Cheggour, B. Devreese, G. Van Driessche, B. Joris, J. Van Beeumen, J. M. Frère, *Biochem. J.* **1999**, *341*, 147–155; c) C. Bompard-Gilles, H. Remaut, V. Villeret, T. Prangé, L. Fanuel, M. Delmarcelle, B. Joris, J. Frère, J. Van Beeumen, *Structure* **2000**, *8*, 971–980.
[16] a) Z. Zou, M. Nöth, F. Jakob, U. Schwaneberg, *Bioconjugate Chem.* **2020**, *31*, 2476–2481; b) M. Wójcik, K. Szala, R. van Merkerk, W. J. Quax, Y. L. Boersma, *Proteins* **2020**, *88*, 1394–1400; c) K. Strijbis, E. Spooner, H. L. Ploegh, *Traffic* **2012**, *13*, 780–789.
[17] I. Wuethrich, J. G. C. Peeters, A. E. M. Blom, C. S. Theile, Z. Li, E. Spooner, H. L. Ploegh, C. P. Guimaraes, *PLoS One* **2014**, *9*, https://doi.org/10.1371/journal.pone.0109883.
[18] D. C. F. Monteiro, V. Patel, C. P. Bartlett, S. Nozaki, T. D. Grant, J. A. Gowdy, G. S. Thompson, A. P. Kalverda, E. H. Snell, H. Niki, A. R. Pearson, M. E. Webb, *Chem. Biol.* **2015**, *22*, 492–503.
[19] Z. Wu, H. Hong, X. Zhao, X. Wang, *Bioresour. Bioprocess.* **2017**, *4*, 13.
[20] A. K. Kenworthy, S. S. Schmieder, K. Raghunathan, A. Tiwari, T. Wang, C. V. Kelly, W. I. Lencer, *Toxin Rev.* **2021**, *13*, 543.
[21] J. P. Dolan, D. C. Machin, S. Dedola, R. A. Field, M. E. Webb, W. B. Turnbull, *Front. Chem.* **2022**, *10*, 958272.
[22] Y. Xia, F. Li, X. Zhang, S. Balamkundu, F. Tang, S. Hu, J. Lescar, J. P. Tam, C.-F. Liu, *J. Am. Chem. Soc.* **2023**, *145*, 6838–6844.
[23] C. P. Guimaraes, M. D. Witte, C. S. Theile, G. Bozkurt, L. Kundrat, A. E. Blom, H. L. Ploegh, *Nat. Protoc.* **2013**, *8*, 1787–1799.
[24] S. Nozaki, M. E. Webb, H. Niki, *MicrobiologyOpen* **2012**, *1*, 298–310.
[25] M. R. Balmforth, J. Haigh, V. Kumar, W. Dai, C. Tiede, D. C. Tomlinson, J. Deuchars, M. E. Webb, W. B. Turnbull, *Bioconjugate Chem.* **2021**, *32*, 2205–2212.

Manuscript received: July 28, 2023
Accepted manuscript online: December 10, 2023
Version of record online: January 11, 2024