Biochemistry and Chemical Biology

Enzymatic protein fusions with 100% product yield

Adrian CD Fuchs

Department of Protein Evolution, Max Planck Institute for Biology, Tübingen, Germany

Reviewed Preprint
v2 • February 27, 2025
Revised by authors

Reviewed Preprint
v1 • December 19, 2024

eLife Assessment

This revision of **important** work is a versatile addition to the chemical protein modifications and bioconjugation toolbox in synthetic biology. The technology developed cleverly uses Connectase to irreversibly fuse proteins of interest together so they can be studied in their native context, with **compelling** well-controlled data showing the technique works for various protein partners. This work will help multiple fields to explore multi-function constructs in basic synthetic biology. This work will also be of interest to those studying fusion oncoproteins commonly expressed in various human pathologies.

https://doi.org/10.7554/eLife.102765.2.sa3

Abstract

Summary

The protein ligase Connectase can be used to fuse proteins to small molecules, solid carriers, or other proteins. Compared to other protein ligases, it offers greater substrate specificity, higher catalytic efficiency, and catalyzes no side reactions. However, its reaction is reversible, resulting in only 50% fusion product from two equally abundant educts. Here, we present a simple method to reliably obtain 100% fusion product in 1:1 conjugation reactions. This method is efficient for protein-protein or protein-peptide fusions at the N-or C-termini. It enables the generation of defined and completely labeled antibody conjugates with one fusion partner on each chain. The reaction requires short incubation times with small amounts of enzyme and is effective even at low substrate concentrations and at low temperatures. With these characteristics, it presents a valuable new tool for bioengineering.

Introduction

Protein conjugations are used to fuse fluorophores to proteins and to immobilize proteins on beads, microplates, electron microscopy grids, or surface plasmon resonance chips. They enable the generation of antibody conjugates, of segmentally labeled proteins for nuclear magnetic resonance measurements, and the fusion of proteins to the cell surface. Various methods are employed for these applications, often with an excess of one conjugation partner (e.g., 3 - 10

equivalents). Each of them comes with its own advantages and disadvantages. N-Hydroxysuccinimide (NHS) labeling of lysine residues or maleimide labeling of cysteine residues, are easy to use, but can be unspecific[1]. Click chemistry is specific, but requires the introduction of non-biological chemical groups in the fusion partners[2,3]. Split domain methods, such as SpyTag-SpyCatcher, are also specific, but introduce long sequences between the fusion partners[4,5]. Split intein methods introduce only a short "ligation scar", but may suffer from solubility problems and varying efficiency for different substrate pairs[6,7]. Enzymatic methods are simple and, in some cases, specific, but the resulting fusions are reversible and therefore incomplete[8,9]. Nonetheless, they find good use, both in industry and in thousands of academic studies.

The most popular enzyme, *Staphylococcus aureus* Sortase A, catalyzes the reversible fusion of substrates **A** and **B** in the form of **A**-LPXTG (X = any amino acid) and G_{1-5}-**B** to yield **A**-LPXTG_{1-5}-**B** (usually with additional linker sequences) [9,10]. Therefore, it shows good specificity for substrate **A**, but is relatively unspecific towards substrate **B**. Here, the glycine may be replaced by lysine side chains or by other amines[11]. Sortase A has a low affinity for its substrates and therefore requires high substrate concentrations for moderate activity (e.g., $K_{M(LPETG)}$ = 7330 µM, $K_{M(GGGGG)}$ = 196 µM[12]; see also [13]). It also catalyzes the irreversible hydrolysis of both educts and products ($k_{cat,\ hydrolysis}$ = 0.086/s; $k_{cat,\ Ligation}$ = 0.28/s)[12]. Some of these characteristics could be improved by the generation of mutant proteins or by the optimization of reaction conditions[14,15]. Nevertheless, these properties still constrain its applicability, meaning that Sortase A is primarily used for protein-peptide fusions at high concentrations, and with an excess of peptide.

Several other enzymes share these basic characteristics with Sortase A. These enzymes belong to the cysteine (Sortase, Butelase, Asparaginyl endopeptidase) or serine proteases (Trypsiligase, Subtiligase), and bind their substrates through a (thio-)ester bond, which can react with H_2O (hydrolysis, irreversible) or the H_2N group of fusion partners (aminolysis, reversible)[8,9]. Recently, however, an enzyme ligase with entirely different characteristics was discovered[16]. This enzyme, Connectase, binds its substrate through an amide bond, which is hydrolysis-resistant. It therefore exclusively catalyzes conjugations with the H_2N groups of fusion partners. In addition, Connectase acts on a longer recognition sequence, leading to higher substrate specificity and higher catalytic efficiency. This enables entirely different applications, such as the specific labeling of proteins with fluorophores within cell extracts, allowing their in-gel detection with significantly higher sensitivity and signal-to-noise ratio compared to Western blots with tag-specific antibodies [17].

Yet just like other enzyme ligases, Connectase catalyzes a reversible reaction[16]. Connectase from *M. mazei*, for example, binds substrates with the sequence **A**-ELASKDPGAFDADPLVVEI (**Figure 1**, step 1). It then forms a covalent intermediate, **A**-ELASKD-Connectase, with the N-terminal part of this sequence, and cleaves off the C-terminal peptide PGAFDADPLVVEI (**Figure 1**, steps 2-3). This reaction works both ways, meaning that PGAFDADPLVVEI can react with **A**-ELASKD-Connectase to restore Connectase and its substrate, **A**-ELASKDPGAFDADPLVVEI. However, when a second substrate **B** in the form of PGAFDADPLVVEI-**B** is added to the reaction, it can be used instead of the peptide PGAFDADPLVVEI to form the fusion product **A**-ELASKDPGAFDADPLVVEI-**B** (**Figure 1**, steps 4-5).

When using equimolar quantities of educts **A** and **B**, Connectase catalyzes an equilibrium of approximately 50% fusion product **A**-**B** and 50% educts. This is because the same amounts of PGAFDADPLVVEI peptide byproduct and PGAFDADPLVVEI-**B** educt compete for the **A**-ELASKD-Connectase intermediate (**Figure 1**, step 3). Consequently, 100% fusion product can be obtained by removing or inactivating the peptide byproduct, for example by specific enzymatic proteolysis. However, the employed protease would have to act only on the PGAFDADPLVVEI peptide and not

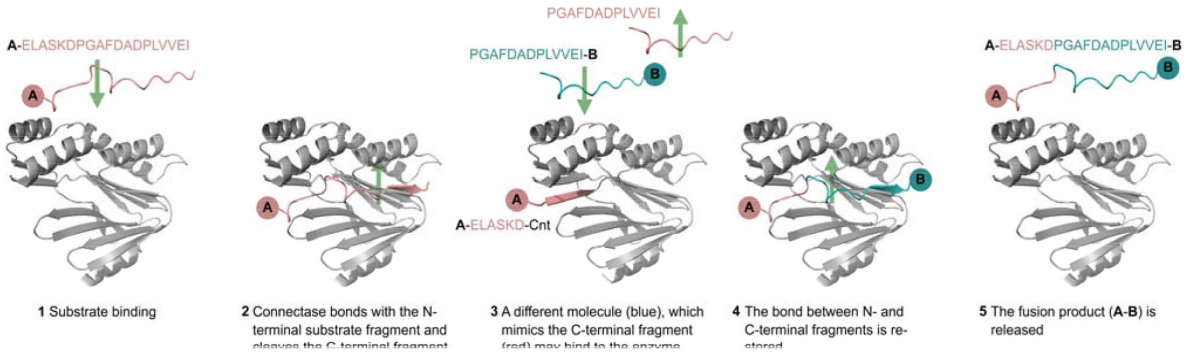


Figure 1

The Connectase reaction mechanism.

The structures were predicted with Alphafold 2. For simplicity, they are shown mirror-inverted, so that the Connectase binding channel is visible, and the recognition sequence can be read from left (N-terminus) to right (C-terminus). **A** and **B** symbolize peptides or proteins.

on the PGAFDADPLVVEI-**B** educt. In the following paper, we present a solution to this problem and describe a simple method to obtain 100% fusion product from equally concentrated educts in short time and with small amounts of enzyme.

Results

Modification of the Connectase recognition sequence

To facilitate the removal of peptide byproduct from Connectase reactions (see introduction), we studied ways to alter the Connectase recognition sequence. For this, we used Ubiquitin (Ub) with a C-terminal Connectase recognition sequence followed by a Streptavidin tag (**Figure 2**, Ub-Strep). As a second reaction substrate, we used peptides derived from the Connectase recognition sequence (XGAFDADPLVVEI, where X represents any of the 20 amino acids). The conjugation of these substrates results in a shorter Ubiquitin product (**Figure 2**, Ub-Peptide), which lacks the Streptavidin tag. Therefore, the conjugation rate with the different peptides could be determined by monitoring Ub-Peptide formation in SDS-PAGE time course analyses (**Figure 2**).

The experiment with the original recognition sequence peptide (X = Proline) resulted in the rapid formation of an equilibrium with ~0.5 equivalents (eq.) Ub-Strep and ~0.5 eq. Ub-Peptide, corresponding to ~50% product yield from equally abundant educts (1 eq. each; see introduction). While proline substitutions with X = D, E, F, G, H, I, K, L, M, N, R, T, W, or Y drastically reduced ligation rates, substitutions with S, C, and A resulted in moderate or high ligation rates. This was surprising, because the KDPGA sequence is highly conserved in the physiological Connectase target, mtrA (methyltransferase A)[16]. Due to its unique structure and chemistry, proline was previously considered essential for the reaction. The results show that other amino acids, which possess a β-carbon (present in all amino acids except G) but lack a γ-carbon (absent in G, S, C, A), can also be used in this position.

Discrimination between different recognition sequences

This unexpected finding allowed us to design reactions, in which educts and peptide byproducts differ in their N-terminal amino acid. This discrimination criterion could then be used to specifically inactivate the X_1GAFDADPLVVEI peptide byproduct (PGAFDADPLVVEI-Strep in **Figure 2**) without affecting the X_2GAFDADPLVVEI-**B** educt (peptide without **B** in **Figure 2**). Such selective inactivation can be achieved through methods that specifically target X_1 (e.g., A, C, S, P, etc.) while leaving X_2 ($\neq X_1$) unmodified. We hypothesized that for example an N-acetyltransferase, which acetylates the N-terminal alanine on the peptide byproduct (X_1 = A), while leaving the N-terminal proline on educts (X_2 = P) unmodified, might be employed to specifically inactivate the undesired byproduct[18]. Another possibility is to use chemicals, which form ring structures with the amino and sulfhydryl-groups of N-terminal cysteines (X_1 = C, X_2 = A / P)[19]. Finally, it is possible to use aminopeptidases, which act exclusively on the peptide byproduct X_1. Many of these enzymes have no absolute specificity for just one amino acid [20]. Proline residues, however, are structurally distinct and not modified by many promiscuous enzymes, but instead by a set of proline-specific enzymes[21].

Based on these considerations, we decided to search for a proline aminopeptidase[22], which removes the N-terminal proline from PGAFDADPLVVEI sequences with suitable efficiency, but is inactive towards all other residues (including X_2 = A). Literature research identified *Bacillus coagulans* proline aminopeptidase (BcPAP) as a candidate. This enzyme had been shown to cleave N-terminal proline from peptides consisting of 2 - 4 amino acids, while remaining inactive towards other N-terminal amino acids[23,24]. We could produce it as a soluble monomer (33 kDa) in *E. coli* (>40 mg from 1 L culture) and tested its suitability for shifting the Connectase reaction equilibrium.

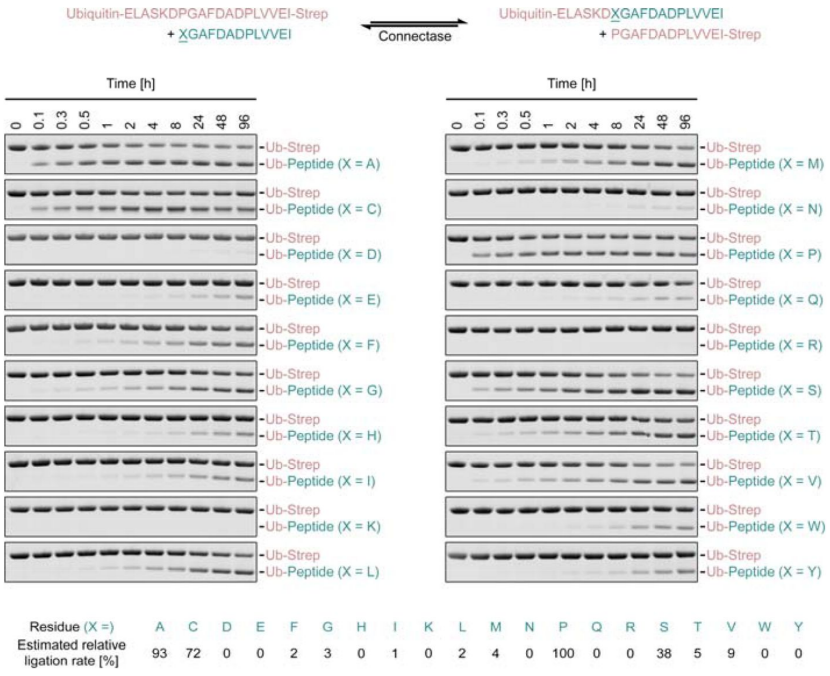


Figure 2

Mutagenic analysis of the Connectase recognition sequence.

A ubiquitin substrate with a C-terminal Connectase recognition sequence followed by a Streptavidin-tag (Ub-Strep) is fused to peptides consisting of N-terminal Connectase recognition sequence variants (top scheme). These peptides differ in the first amino acid, where proline was replaced by each of the 19 other standard proteinogenic amino acids. An SDS-PAGE time course analysis of each reaction (1 eq. Ub-Strep, 1 eq. Peptide, 0.01 eq. Connectase, 22°C) shows the gradual emergence of the fusion product, Ub-Peptide. Based on densitometric analyses, the reaction rate with the different substrates was estimated (lower panel; an exact determination is not possible due to the reversibility of the reaction (see method section)) and normalized to the highest rate (X = P, 100%). The peptide substrates (XGADADPLVVEI) and byproducts (PGAFDADPLVVEI-Strep; top panel) are not visible on the gels.

A method for complete protein-protein fusions

We tested the effect of BcPAP in ligation reactions with **A**-ELASKDPGAFDADPLVVEI (**A** = LysS (Lysine-tRNA ligase), GST (Glutathione-S-Transferase), Ub (Ubiquitin)) and AGAFDADPLVVEI-**B** (**B** = MBP (Maltose Binding Protein), Ub, - (just the peptide)) substrates, which cover a range of molecular weights (1 - 61 kDa). The reactions were performed at room temperature (22°C) in a neutral buffer (pH 7.0), with moderate salt concentrations (150 mM NaCl, 50 mM KCl), 100 µM of each substrate (A and B), as well as 0.033 eq. Connectase, and 0.066 eq. BcPAP. They were separated by SDS-PAGE, stained with Coomassie G-250, and imaged with a fluorescence scanner (Excitation 685 nm, Emission 725 nm). This allowed the densitometric quantification of the resulting protein bands with good accuracy.

In each case (**Figure 3A-C**, **Figure S2A**), we observed 98 - 100 % conversion of the less abundant substrate without any reaction side products. In protein-protein ligations (**Figure 3A-B**, **Figure S2A**), ~90% fusion product was obtained after one hour of incubation, and ~95% after two hours. Protein-peptide ligations (**Figure 3C**) proceeded even faster. The reaction rate was approximately four times slower at low substrate concentrations (10 µM (**Figure 3D**) instead of 100 µM (**Figure 3A**)) and about eight times slower at low temperatures (10°C (**Figure 3E**) instead of 22°C (**Figure 3A**)), but still resulted in complete protein-protein fusions.

We generated the AGAFDADPLVVEI-**B** substrates for these experiments by TEV protease cleavage of MENLYFQ|AGAFDADPLVVEI-**B** precursors. We chose this approach to avoid the potential acetylation of N-terminal alanine residues during expression of (M)AGAFDADPLVVEI-**B** substrates (methionine removal by methionine aminopeptidase). The cleavage is efficient as the TEV protease recognition sequence is exposed N-terminally of the unstructured Connectase recognition sequence. Consequently, cleavage and conjugation can be performed in parallel with small amounts of enzyme (**Figure 3F**). Another way to prevent N-terminal acetylation is to add an extra N-terminal proline (i.e., P|AGAFDADPLVVEI-**B**), which is removed during the reaction by BcPAP (**Figure S2B**).

To substantiate these findings, we conducted a liquid chromatography mass spectrometry (LC-MS) analysis. The signal intensities in these experiments allow no protein quantifications, as smaller molecules are often measured with more intense signals. Nevertheless, we report relative intensities in this paper as a qualitative measure. For a 1:1 Ub-MBP conjugation (**Figure S3**), they amounted to 0.25% (Ub), 0.13% (MBP), and 99.6% (Ub-MBP). The N-terminal proline was almost completely (99.8%) removed from the peptide byproduct, (P)GAFDAPLVVEI. Similar to all other LC-MS tested molecules in this study (i.e. Connectase, BcPAP, MBP, LysS, αHER2 LC/HC, Ubiquitin; **Figures 4** and **5**), it was not further processed on the N-terminus. This supports the finding[20,23,24] that BcPAP acts exclusively on N-terminal proline residues.

Finally, we tested the effect of serine-, cysteine-, or metalloprotease inhibitors on the reaction. Connectase is not a protease[16] and therefore unaffected by these substances (**Figure S4**). However, the equilibrium shift associated with BcPAP activity could be suppressed with the serine protease inhibitors PMSF and AEBSF (**Figure S4**). These results contrast with previous studies, where BcPAP was found to be more susceptible to cysteine protease inhibitors[24]. They are, however, consistent with the classification of BcPAP as a serine protease[25].

Homogeneous antibody conjugates

Antibodies are the most relevant protein conjugation target[26]. Many applications require their conjugation to spacious payload molecules (e.g., horseradish peroxidase) and/or to a defined number of molecules. This number, also known as drug : antibody ratio, is used as a benchmark for several specialized techniques. Of these, formyl-glycine insertion[27] (Catalent), sugar engineering[28] (Mersana), cysteine engineering[29] (Genentech), the introduction of unnatural

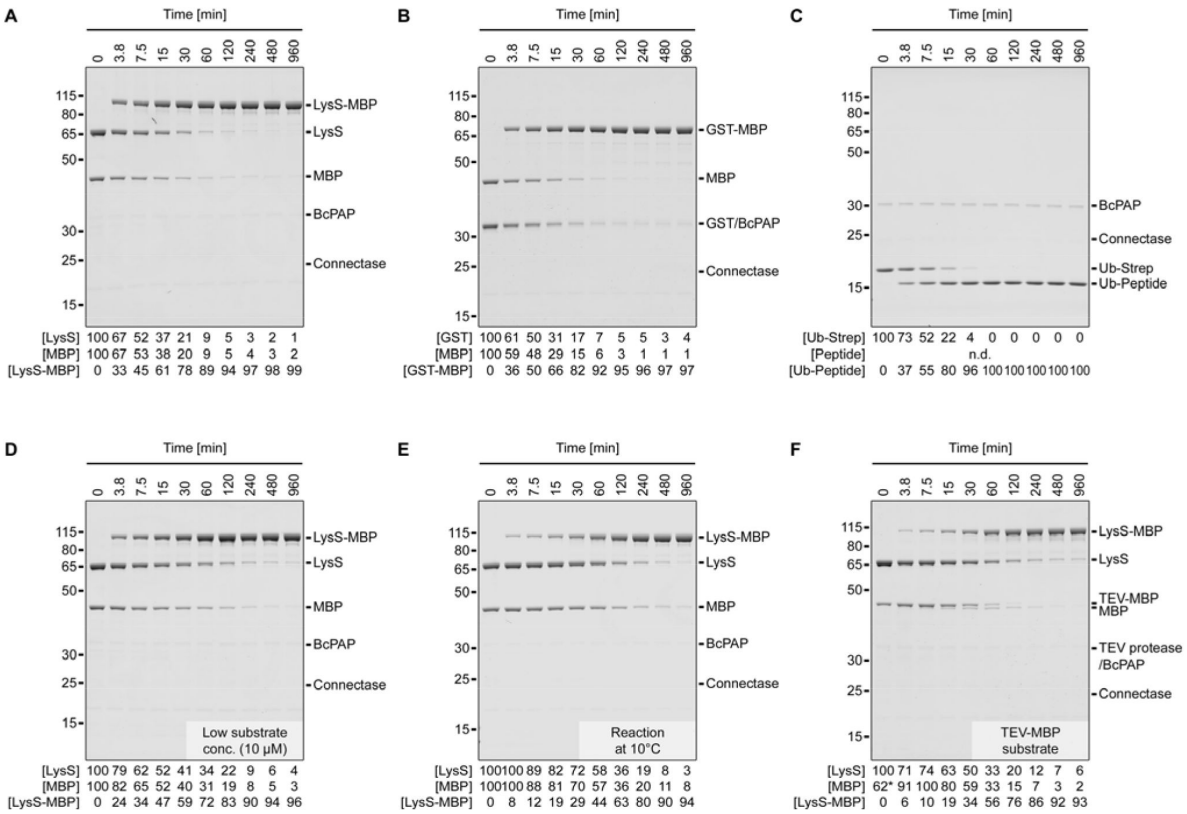


Figure 3

Complete protein-protein ligations.

Shown are SDS-PAGE time course analyses of ligation reactions using Lysine-tRNA ligase (LysS) and Maltose binding protein (MBP; A, D, E, F), Glutathione-S-Transferase (GST) and MBP (B), or Ubiquitin with Streptavidin tag (Ub-Strep) and AGAFDADPLVVEI peptide (C) as substrates. Each reaction was performed with 1 eq. N-terminal fusion partner (LysS, GST or Ubiquitin; C-terminal ELASKDPGAFDADPLVVEI sequence), 1 eq. C-terminal fusion partner (MBP or peptide; N-terminal AGAFDADPLVVEI sequence), 0.033 eq. Connectase, and 0.066 eq. BcPAP. The substrate concentration was 100 µM (except for D: 10 µM) and the incubation temperature was 22°C (except for E: 10°C). In experiment F, an MBP protein with an additional N-terminal TEV protease recognition sequence (MENLYFQ|AGAFDADPLVVEI-MBP) was used and TEV protease (0.01 eq.) was added to the reaction. A densitometric analysis of the protein bands is shown below each experiment. For the substrates, the values reflect the substrate band density relative to the substrate band in the control sample (0 min); for the products, the values reflect the product band density relative to the total band density (substrates + products). The exact values can be found in Dataset S2. They are also visualized in **Figure S1**.

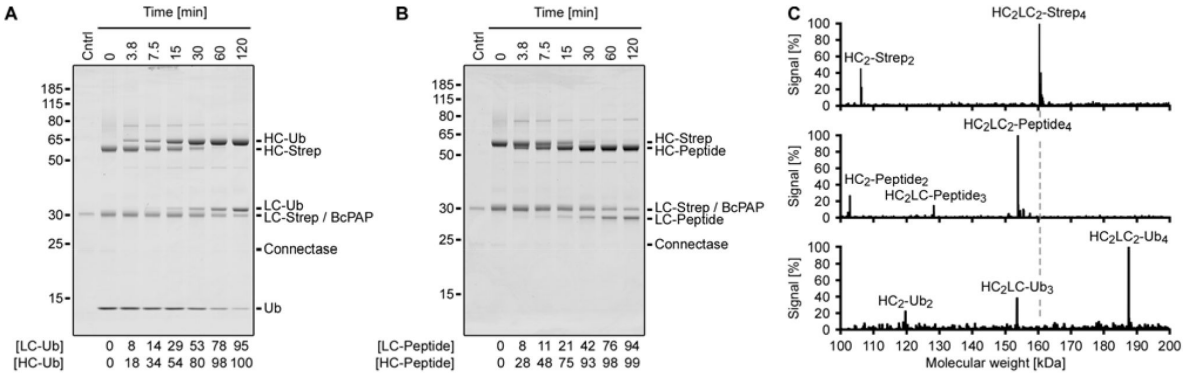


Figure 4

Antibody conjugation.

Shown are SDS-PAGE time course (A, B) and LC-MS analyses (C) of αHER2 (human epidermal growth factor receptor 2) antibody conjugations. The αHER2 heavy (HC) and light chains (LC) were produced with a C-terminal Connectase recognition sequence and a Streptavidin tag (HC-Strep, LC-Strep). In the reactions (25 μM αHER2 (100 μM subunits), 0.033 eq. Connectase, 0.066 eq. BcPAP, 22°C), the Streptavidin tag is replaced by Ubiquitin (A; 1 eq.) or a shorter peptide (B; 1 eq.). A densitometric quantification of the product bands relative to the educt bands is shown below the gels. For the calculation, the BcPAP density in the control lane (Cntrl) was subtracted from the combined LC-Strep/BcPAP band. The LC-MS analyses (C) show the assemblies in the unconjugated antibody sample (top panel) and a shift of the detected masses, consistent with a near-complete conjugation to peptide (middle panel) or ubiquitin (lower panel). All detected masses, their abundances, and assignments can be found in the Dataset S2.

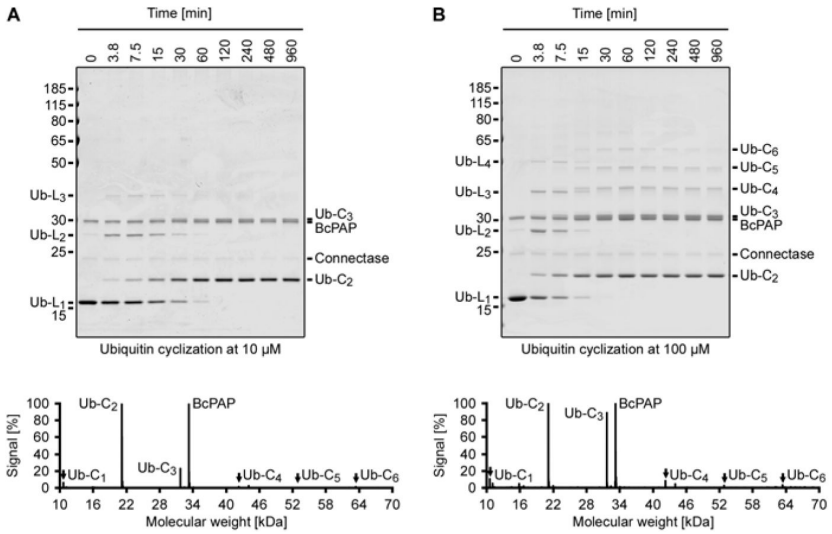


Figure 5

Protein cyclization.

Shown are SDS-PAGE time course analyses of a Ubiquitin cyclization reaction. The employed Ubiquitin substrate was produced with both an N-terminal (AGAFADPLVVEI) and a C-terminal (ELASKDPGAFDADPLVVEI) Connectase recognition sequence. This allows the formation of linear (L_1 - L_4, formed by 1 - 4 Ubiquitin proteins) polymers, which are observed in the early stages of the time course. The N-terminus of a given polymer can be fused to its C-terminus, resulting in cyclic assemblies (C_2 - C_6, formed by 2 - 6 Ubiquitin proteins), which present the end product of the reaction. A lower substrate concentration (A, 10 µM) results in smaller assemblies, and a higher substrate concentration results in larger assemblies (B, 100 µM). The assignment of the gel bands is consistent with LC-MS data (below the gels). The plots were normalized to the most intense Ubiquitin signal (Ub-C_2); the BcPAP peaks are more intense (>100%), despite its relatively low abundance. The molecular masses of the ubiquitin assemblies are 13 kDa (L_1), 24 kDa (L_2), 34 kDa (L_3), 45 kDa (L_4), 21 kDa (C_2), 32 kDa (C_3), 43 kDa (C_4), 53 kDa (C_5), and 64 kDa (C_6).

amino acids[30] (Sutro), and Sortase-mediated conjugations[31] (NBE) all find commercial use. Each of these approaches has its advantages and disadvantages. One problem is that even a relatively high conjugation ratio of 90% on each antibody chain leads to a desired drug : antibody ratio of 4 in only 66% of all HC_2LC_2 antibodies, as each antibody chain is modified separately. The completely labeled antibodies are hard to separate from the partially labeled antibodies and side reactions can further complicate the process.

To test the established method for antibody conjugation, we added the Connectase recognition sequence plus an additional C-terminal Streptavidin tag to the C-termini of αHER2 light and heavy chains. HEK293 cells were transiently transfected with plasmids encoding for these proteins, and the culture medium was tested for the exported antibodies with the in-gel fluorescence method[17,32]. In this western blot alternative, Connectase is used to fuse far-red fluorophores to target proteins, which can subsequently be detected on the SDS-gel with a fluorescence imager (**Figure S5**). From this, we expected a concentration of ~2 µg antibody per ml of culture medium – a value that could be confirmed after protein purification with a Streptavidin column.

The purified antibodies were labeled with either AGAFDADPLVVEI peptide, AGAFDADPLVVEI-Ubiquitin (1 eq. per antibody subunit; reaction conditions as in **Figure 3**), or AGAFDADPLVVEIK(−Biotin) (**Figure S6**). SDS-PAGE time course analyses show almost complete labeling reactions after an incubation time of two hours (**Figure 4A-B**). The heavy chain is labelled faster than the light chain, possibly because the Connectase recognition sequence is more accessible in this case. The conjugation products remained stable for several days in blood plasma (**Figure S7**).

The reactions were also analyzed by LC-MS. For this, the antibody carbohydrate side chains were removed enzymatically to reduce sample complexity[33]. The masses determined in the unconjugated antibody sample were consistent with assemblies of completely intact antibody subunits, except for the C-terminal Strep-tag lysine, which was missing in most subunits. Of the assigned masses, HC_2LC_2 assemblies constituted 69% of the signal, followed by HC_2 (30%), and HC_2LC_1 (0.4%) assemblies (**Figure 4C**, top panel).

Next, a 1:1 antibody-peptide conjugation reaction was analyzed with the same method (**Figure 4C**, middle panel). Here, the HC_2LC_1 assembly was detected only in completely conjugated form, as HC_2LC_1-$peptide_3$ (100%). Similarly, HC_2 was completely conjugated to HC_2-$peptide_2$ (100%). HC_2LC_2 was found to be conjugated to either two (2%), three (15%), or four (83%) peptides. Taken together, this suggests a total labeling efficiency of 97%, in line with the results shown in **Figure 4B**.

Finally, a 1:1 antibody-ubiquitin conjugation was analyzed (**Figure 4C**, lower panel). Here, the HC_2LC_1 assembly was more abundant and the HC_2 assembly less abundant compared to the unconjugated sample. Again, HC_2 and HC_2LC_1 were detected only in completely conjugated form, as HC_2-Ub_2 (100%) and HC_2LC_1-Ub_3 (100%). HC_2LC_2 was found to be conjugated to either one (2%), two (1%), three (2%), or four (95%) ubiquitin molecules, suggesting a total labeling efficiency of 98%.

These results demonstrate that the method is effective for the rapid generation of near-homogeneous antibody conjugates. The conjugation ratio could potentially be further increased with a small excess of the conjugation partner (here: Ubiquitin or peptide). The desired products can be purified in subsequent steps, for example with Streptavidin columns to remove unconjugated antibodies, or with size exclusion columns, to separate HC_2LC_2 from other assemblies.

Protein cyclization and polymerization

Protein cyclization is used to enhance protein stability[34], while protein polymerization can be useful for engineering biomaterials or affine binders[35]. Both results may be achieved with proteins carrying both N- and C-terminal Connectase recognition sequences. Circularization may be favored, when the protein is present at a low concentration and when its N- and C-termini are in close proximity. Polymerization might be favored for rod-shaped proteins with distant N- and C-termini at high concentrations.

To test these ideas, we used Ubiquitin (AGAFDADPLVVEI-Ub-ELASKDPGAFDADPLVVEI) as a small globular test substrate at low concentrations (10 µM, **Figure 5A**). At the start of the reaction (0 – 7.5 min), we observed its polymerization to linear chains of two (Ub-L_2) or three (Ub-L_3) protomers. These linear assemblies gradually disappeared (7.5 – 120 min), as their N- and C-termini were fused. The resulting circular assemblies (80% Ub-C_2 and 20% Ub-C_3, according to SDS-PAGE densitometry) presented the stable end-products of the reaction.

When the reaction was performed with ten-fold higher substrate concentrations (100 µM, **Figure 5B**), longer linear ubiquitin polymers were initially formed, which eventually gave rise to larger circular assemblies (37% Ub-C_2, 45% Ub-C_3, 9% Ub-C_4, 6% Ub-C_5, and 2% Ub-C_6). Single ubiquitin molecules were not efficiently circularized in either experiment, suggesting that the N- and C-termini in one ubiquitin molecule are too far apart to be connected, and that additional linker sequences are needed to favor this assembly.

Taken together, the results show that the presented conjugation method enables efficient protein circularizations and that the yields of specific assemblies can be controlled by modifying the reaction conditions (e.g., substrate concentration) and the substrate design (e.g., linker length). As with other cyclization/polymerization methods[36], these parameters need to be tested for each individual substrate.

Discussion

In this paper, we have shown that the Connectase recognition sequence can be altered, so that the peptide reaction byproduct can be specifically inactivated by a proline aminopeptidase. This method enables the fast, simple, specific, and complete 1:1 fusion of proteins and/or peptides.

Instead of just fusing two molecules, as for example in chemical ligations, the method replaces one sequence (e.g. PGAFDADPLVVEI-Strep in **Figure 4**) with the fusion partner. This has several advantages. For example, an affinity tag used for substrate purification can be removed by the conjugation reaction. As the affinity tag remains on unconjugated substrates, it can subsequently be used to remove the peptide byproduct and any unconjugated substrates (e.g., an antibody conjugated to only three molecules instead of four). When the same affinity tag is added to Connectase and BcPAP, this setup enables the complete purification of homogeneous fusion product in a single step (**Figure 6**).

Another consequence of this replacement mechanism is that complete fusions may be reversed on demand. This is not possible with chemical, split intein, or split domain ligation methods. Yet, the addition of a new X_3GAFDADPLVVEI-**C** fusion partner to an existing **A**-ELASKDX_2GAFDADPLVVEI-**B** fusion product allows the exchange of molecules **B** and **C**. To make this second conjugation complete, a new method for the specific removal of X_2 is needed, just like BcPAP was used to remove X_1 = Pro in this study. This new method could involve another aminopeptidase[20], a specific chemical modification[19], or an N-acetyltransferase[18] (see the section “Discrimination

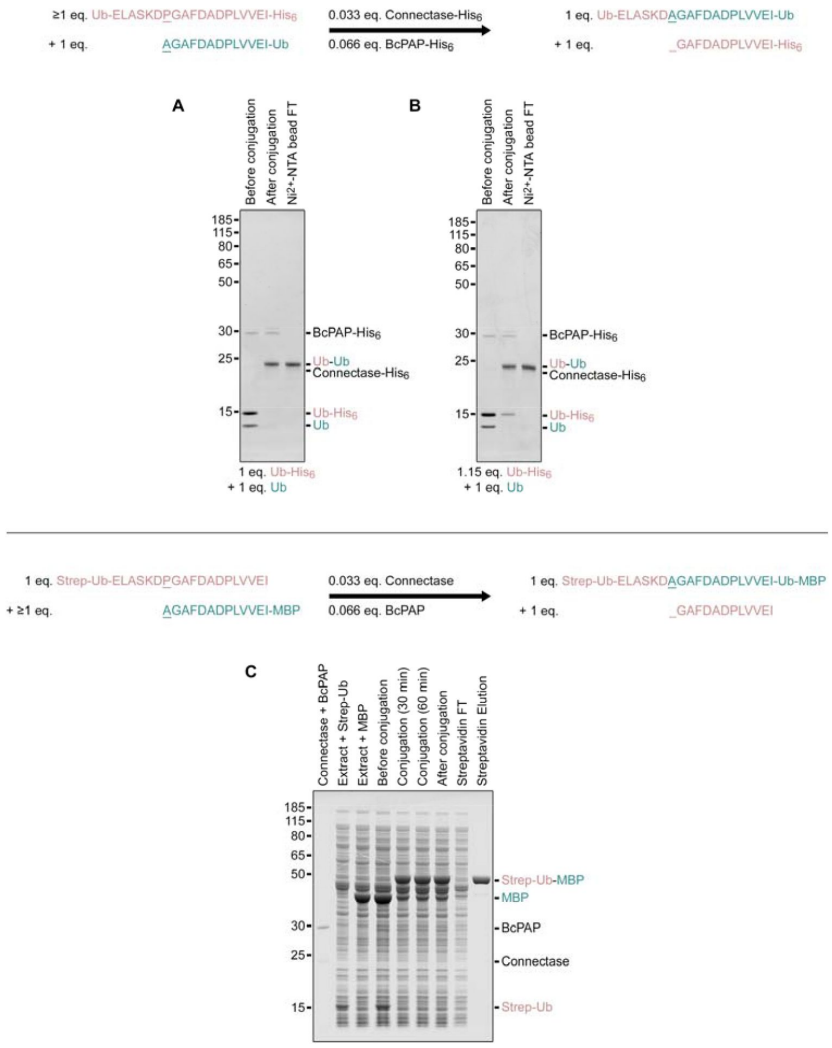


Figure 6

Two exemplary strategies for purifying protein conjugates.

(A, B) The first strategy uses purified educts and couples the conjugation reaction with affinity tag removal (top scheme). This allows the separation of the untagged product (Ub-Ub conjugate) from tagged enzymes (Connetase-His_6, BcPAP-His_6), side products (GAFDADPLVVEI-His_6 peptide), and the first educt (Ub-His_6) by collecting the unbound fraction (flow through (FT)) in an affinity chromatography step. The second untagged educt (AGADADPLVVEI-Ub) must be entirely converted. This can be achieved with sufficient incubation times in equimolar reactions (A) or by using a small excess of reaction partner (B).
(C) The second strategy can be used for expressed proteins in cell extracts and introduces an affinity tag in the reaction product (top scheme). It necessitates the complete conversion of the first affinity tagged educt (Strep-Ub).
This is most conveniently achieved by using the second untagged educt (MBP) in excess. Note that proteins with N-terminal alanine residues, such as AGAFDADPLVVEI-MBP, may be partially N-acetylated in some expression systems. In these cases, a higher excess (e.g., more extract with MBP) must be used, as acetylated substrates are not converted by Connectase.
Strategies to avoid N-acetylation are described in **Figures 3** and **S2**.
Aside from these two strategies, alternative methods such as tandem affinity purification or size exclusion chromatography can also be employed.

between different recognition sequences"). The resulting re-conjugation would enable advanced applications, such as the immobilization of a protein on a surface and its release on demand, or the visualization of existing conjugation products by fluorophore (de-) coupling.

These considerations illustrate the versatility of Connectase as a bioengineering tool. As more applications with this enzyme are developed, synergies will emerge. For example, in this paper, the same Connectase recognition sequence could be used to monitor antibody expression levels in a western blot-like application (**Figure S5**) and for the subsequent antibody conjugation. In future, other Connectase applications, such as protein purification or microplate-based protein quantification, may offer additional functionality for the recognition sequence.

During the preparation of this study, a somewhat related method was published for the enzyme Sortase A[37]. The authors fused various sequences, such as **A**-LPETGAHHHHH and GVGSKYG for **A**-LPETGVGSKYG generation (C-terminal labeling) or YALPETGG and GVGK-**B** for YALPETGVGK-**B** generation (N-terminal labeling). The peptide byproducts (i.e., GAHHHHHH or GG), were digested with an aminopeptidase. This aminopeptidase showed a preference for GG or GA over GV and therefore displayed a lower activity towards GVGSKYG or GVGK-**B** substrates. The method significantly increased the obtained fusion product yields and therefore presents an important development for Sortase-mediated ligations. Despite these similarities, there are important differences compared to the method presented here. As described in the introduction, Sortase irreversibly hydrolyzes both educts and products, shows low specificity for the C-terminal fusion partner, and has low catalytic efficiency[8,9]. As a result, the method is most efficient for C-terminal protein conjugations with small peptides at high concentrations[37]. Protein-protein conjugations are relatively inefficient, as they result in high quantities of hydrolysis side products and remain largely incomplete, even after several days of incubation with high enzyme quantities.

The main disadvantage of Connectase lies in its relatively long recognition sequence. The substrates require a 13 amino acid recognition sequence on the N-terminus and a 19 amino acid sequence on the C-terminus (**Figure S8**). For protein-protein conjugations, we employed additional linker sequences of one (N-terminal R) and five (C-terminal AAAGA) amino acids. This design resulted in efficient conjugations for each substrate tested so far, including 44 constructs derived from 19 different proteins[16,17,32]. Therefore, the employed linkers can generally be regarded as sufficient, even in cases of poor steric accessibility of the substrate termini. Taken together, this results in a final conjugation "scar" of 19 - 25 amino acids between the fusion partners. This is comparable to the SpyTag / KTag system (23 amino acids plus linker sequences)[35], but longer than the sequences typically employed for Sortase-mediated fusions (7 - 9 amino acids, usually with additional $(GGGGS)_x$ linkers between the protein(s) and the ligation site[38]). Long linker sequences can potentially interfere with protein folding, and they can be susceptible to proteases, or proof immunogenic. Though these problems have not been encountered for Connectase so far, it is an attractive option to engineer Connectase variants with a shorter recognition sequence in the future. For now, the method described here is primarily useful in applications, where this longer "scar" can be tolerated. In these cases, however, the combination of simplicity, substrate specificity, efficiency, and completeness, along with the possibility to use either biologically or chemically produced substrates, and the absence of side reactions, is unmatched by other chemical or enzymatic methods.

Methods

Cloning, Expression, and Purification

The sequences of all proteins and peptides used in this study are listed in Dataset S1. The peptide for the in-gel fluorescence assay was synthesized by Intavis, while all other peptides were synthesized by Genecust. Genes were synthesized by Biocat, and cloned into the pET30b(+) vector

(restriction sites: NdeI, XhoI) for expression in *E. coli* or the pcDNA3.1 vector (restriction sites: HindIII, XhoI) for expression in HEK293 cells.

For recombinant expression in *E. coli*, BL21 gold cells were transfected with the respective plasmids and grown in lysogeny broth medium with 50 µg/l kanamycin at 22°C. Protein expression was induced at an optical density of 0.4 at 600 nm with 500 µM isopropyl-β-D-thiogalactoside. Cells expressing soluble proteins were harvested after 16 h, resuspended in buffer (100 mM Tris-HCl, 1x c0mplete EDTA-free protease inhibitor cocktail (Roche; no inhibitor was added to cells expressing BcPAP), 0.02 g/l DNAse, pH 8.0), lysed by French press, and cleared from cell debris by ultracentrifugation (120000 *g*, 45 min, 4°C).

For recombinant expression of αHER2 antibodies, HEK293 cells were cultured at 37°C in ten 75 cm^2 flasks with Dulbecco's Modified Eagle Medium (DMEM) supplemented with fetal calf serum. At 70% confluency, they were transfected with plasmids encoding for αHER2 light and heavy chains using Lipofectamine 2000 (Thermo), according to the manufacturer's instructions (47 µl Lipofectamine, 55 µg of each plasmid). The cells were grown for ten days without splitting, and the medium was exchanged daily. The medium samples were used for αHER2 detection by in-gel fluorescence (**Figure S5**, described below), then pooled, centrifuged (6000 g, 10 min, 4°C), and filtered (0.45 µm) before protein purification.

For protein purification, His_6-tagged proteins (all proteins except for antibody subunits (**Figure 4**, **S5**, and **S7**) and Strep-tagged Ubiquitin constructs (**Figures 5**-**6**)) were applied to HisTrap HP columns (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 20 - 250 mM imidazole). Strep-tagged proteins (the antibody subunits and the Ubiquitin constructs) were instead purified with StrepTrap XT columns (1.8 mM KH_2PO_4, 10 mM Na_2HPO_4, 2.7 mM KCl, 138 mM NaCl, 0 - 50 mM Biotin, pH 7.4). After this initial purification step, proteins with N-terminal TEV recognition sequences (MBP (**Figures 3** and **S3**), Ubiquitin (**Figures S2** and **4**), and Ubiquitin for cyclization (**Figure 5**)) were incubated with TEV protease at a 1:100 molar ratio. The reaction was performed overnight in dialysis tubes (dialysis buffer: 20 mM Tris-HCl pH 8.0, 250 mM NaCl) at 4°C. The processed proteins were separated from His6-tagged TEV protease, N-terminal fragments (MHHHHHHENLYFQ), and residual unprocessed proteins by another purification step. For this, the reactions were applied a second time to HisTrap HP columns (as above), and the flow-through was collected. All chromatography steps were performed on an Äkta Purifier FPLC (GE Healthcare) using Unicorn v5.1.0 software. Purified proteins (**Figure S9**) were supplemented with 15% glycerol, flash-frozen in liquid nitrogen, and stored at −80°C.

Biochemical Assays

Unless noted otherwise, all conjugation reactions were performed at 22°C in neutral (pH 7.0) buffer containing 50 mM sodium acetate, 50 mM MES, 50 mM HEPES, 150 mM NaCl, and 50 mM KCl. They were stopped with SDS loading buffer (final concentration: 50 mM Tris-HCl, 2% SDS, 10% glycerol, 25 mM β-mercaptoethanol, 0.01% bromophenol blue, pH 6.8; final protein concentration ~0.1 g/l) and incubated at 90°C for 10 minutes. The samples were separated using mPAGE 12% Bis-Tris gels (Merck; 5 µl loading volume per sample) with MOPS running buffer (50 mM MOPS, 50 mM Tris, 0.1% SDS, 1 mM EDTA; no pH adjustment). The gels were stained with Coomassie blue (25% ethanol, 25% 25% methanol, 10% acetate, 0.25% Coomassie R-250), and subsequently with Coomassie colloidal solution (20% ethanol, 10% ammonium sulfate, 5.8% phosphoric acid, 5% methanol, 0.12% Coomassie G-250). They were destained with 10% acetic acid and imaged with an Azure Sapphire NIR fluorescence scanner (excitation at 685 nm, emission at 725 nm, 25 - 50 µm resolution, Intensity 8, highest scanning speed). Densitometric band quantification was performed with Image Studio Lite 5.2. The bands were manually assigned with the "draw rectangle" tool. Each quantified region encompassed the entire band, bordered by a minimum of three pixels of background for accurate measurement. All obtained values and all unprocessed gels can be found in Dataset S2.

For **Figure 2**, 20 experiments were set up, each with a different XGAFDADPLVVEI peptide (X = any of the 20 amino acids). The reactions contained 20 mM Ub-Strep, 20 mM peptide, and 0.2 mM Connectase. Samples were taken before the addition of Connectase (0 min) and after the indicated times (0.1 - 96 h). After SDS-PAGE and densitometric analyses, the relative reaction rates in each experiment were estimated. An exact determination was not possible because the reaction is reversible, and the emerging peptide side product (PGAFDADPLVVEI-Strep) competes with the assayed peptide (XGAFDADPLVVEI) for the enzyme binding sites. The estimates were made based on the required time to obtain 10%, 20%, and 30% fusion product yield in each reaction.

For **Figure 3** and **Figure S2**, 8 experiments were set up with different substrate pairs (**Figure 3A, D**, E: LysS/MBP; 3B: GST/MBP; 3C: Ub-Strep/AGAFDADPLVVEI peptide; 3F: LysS/MBP before TEV cleavage (see "Purification"); **Figure S2A**: LysS/Ub; S2B: LysS/Pro-Ub; protein sequences are listed in Dataset S1). Each substrate was used at 100 µM (except for **Figure 3D** (10 µM)). The reactions were started by addition of 0.033 molar equivalents (eq.) Connectase and 0.066 eq. BcPAP. For the experiment shown in **Figure 3F**, 0.01 eq. TEV protease was also added. The reactions were incubated at 22°C (except for **Figure 3E** (10°C)). Samples were taken before (0 min) and after the addition of Connectase/BcPAP (3.8 - 960 min). After SDS-PAGE, the substrate band densities relative to the control sample (0 min) were determined. The product band densities were determined relative to the total band density (substrates + products).

For **Figure S4**, Connectase (2 µM) was used without (A) or with BcPAP (7 µM, B). Both solutions were incubated at 22°C for 30 min with the following compounds: buffer (control reaction), 1 mM ZnCl2, "Complete" protease inhibitor mix (Roche, 1 tablet per 25 ml), AEBSF, ALLN, Antipain, Aprotinin, Bestatin, Chymostatin, E-64, EDTA, Leupeptin, Pepstatin, Phosphoramidon, PMSF. The concentrations of the last compounds are unknown to us, as the supplier (G-biosciences, Protease Inhibitor Set, inhibitors used at "2x" concentration) refused to provide them on request. After the incubation, the enzyme-inhibitor mixture was added to an equal volume of conjugation substrates (20 µM Ub-Strep, 20 µM AGAFDADPLVVEI peptide). After 2 h at 22°C, the reactions were analyzed by SDS-PAGE.

For **Figure S5**, the cell culture medium samples taken during αHER2 expression (see above) were analyzed. They were centrifuged (1 min, 10000 g) and 2.5 µl of each supernatant was mixed with 300 fmol reference protein (MBP with a C-terminal Connectase recognition sequence). The mixture was incubated with 1 nM Connectase and 10 nM fluorescent peptide substrate (RELASKDPGAFDADPLVVEISEEGE-Cy5.5) for 20 min at 22°C. The reactions were separated by SDS-PAGE and imaged before and after Coomassie staining. The band densities corresponding to reference protein and αHER light (LC) and heavy chains (HC) were determined. They were used to estimate the expressed αHER2 quantities over time with [αHER2] = reference protein quantities x ([signal HC] + [signal LC] / [signal reference protein]). This estimation can be turned into an exact determination of αHER2 quantities by division with an experimentally determined factor, which describes the relative reactivity and brightness of antibody and reference protein bands. In this case, this factor was close to 1, so that estimated and determined values were nearly identical. The determination and the detailed method are described in [32].

For **Figure 4** and **Figure S6**, 25 µM Strep-tagged αHER2 antibody was mixed with 3.33 µM Connectase, 6.66 µM BcPAP, and either 100 µM AGAFDADPLVVEI-Ubiquitin, 100 µM AGAFDADPLVVEI peptide (**Figure 4**), or 100 µM AGAFDADPLVVEIK(−Biotin) (**Figure S6**). The reactions were incubated at 22°C, and samples were taken after the indicated times (0 - 120 min). After SDS-PAGE, the product band quantities (i.e., LC-Ub, HC-Ub, or LC-peptide, HC-peptide) were determined relative to the educt band quantities (LC-Strep, HC-Strep).

For **Figure S7**, αHER2 antibodies were biotinylated as described above and mixed (1 µM) with human plasma (a gift from Dr. Konnerth; supplemented with EDTA to prevent clotting). The mixture was divided into three tubes and incubated at 4°C, 22°C, or 37°C. Samples were taken after

0, 1, 3, 5, and 7 days and stored at −20°C. For the analysis, 2 µl of each sample was mixed with 33 µl labeling solution (20 nM Connectase, 2 nM PGAFDADPLVVEISEEGE-Cy5.5 peptide) and incubated for 30 min. The samples were then supplemented with 15 µl SDS loading buffer and separated by SDS-PAGE (5 µl per lane). The gel was first analyzed by in-gel fluorescence (see above (**Figure S5**)) and subsequently stained with Coomassie G-250.

For **Figure 5**, a ubiquitin variant with an N-terminal (AGAFDADPLVVEI…) and a C-terminal (ELASKDPGAFDADPLVVEI) Connectase recognition sequence was employed. This substrate was used at a concentration of 10 µM (first experiment) and at a concentration of 100 µM (second experiment). Both reactions were conducted with 0.033 eq. Connectase and 0.066 eq. BcPAP at 22°C. Samples were taken after the indicated times and analyzed by SDS-PAGE.

For **Figure 6**, 100 µM (A) or 115 µM (B) Ubiquitin with C-terminal Connectase recognition sequence was mixed with 100 µM Ubiquitin with N-terminal Connectase recognition sequence, 6.66 µM BcPAP, and 3.33 µM Connectase. The mixtures were incubated overnight (A) or for 2 hours (B), and applied to a Ni^{2+}-NTA column (see purification). Samples were collected at the start of the reaction, after the incubation, and after the chromatography step (flow-through), and analyzed by SDS-PAGE. For **Figure 6C**, Streptavidin-tagged Ubiquitin with N-terminal Connectase recognition sequence and processed MBP (N-terminal AGAFDADPLVVEI sequence, see **Figure 3**) were employed. Both proteins were mixed with *E. coli* cell extract (lane 2, 3), combined (final concentrations: ~1 g/l Ubiquitin (~80 µM), ~ 4 g/l MBP (~100 µM), ~40 g/l *E. coli* protein), incubated for 90 min at 22°C with 6.66 µM BcPAP and 3.33 µM Connectase (lanes 4-6), and applied to a Streptavidin column (see purification). The described samples and the eluted conjugate were analyzed by SDS-PAGE.

For **Figure S8**, Ub-Strep (10 µM) was mixed with RELASKDPGAFDADPLVVEI, ELASKDPGAFDADPLVVEI, or LASKDPGAFDADPLVVEI peptides (10 µM) and 0.25 µM Connectase. Reaction samples were taken after the indicated times (0 - 60 min) and analyzed by SDS-PAGE.

Liquid Chromatography-Mass Spectrometry (LC-MS)

LC-MS analysis was performed at the Natural and Medical Sciences Institute (NMI, Reutlingen, Germany), using established sample preparation and data interpretation protocols[33,39]. Specifically, the samples were prepared as follows:

For **Figure S3**, 100 µM Ub-Strep, 100 µM MBP, 3.33 µM Connectase, and 6.66 µM BcPAP were mixed. The reaction was incubated for 4 h at 22°C and then used for LC-MS analysis.

For **Figure 4**, 25 µM Strep-tagged αHER2 antibody was mixed with 3.33 µM Connectase, 6.66 µM BcPAP, and either 100 µM AGAFDADPLVVEI-Ubiquitin or 100 µM AGAFDADPLVVEI peptide. The reaction was incubated for 4 h at 22°C and then used for LC-MS analysis. Unconjugated αHER2 antibody was used as a control. All samples were deglycosylated with PNGase F (R&D Systems) for 16h at 37°C.

For **Figure 5**, a ubiquitin variant with N- and C-terminal Connectase recognition sequence was used at two different concentrations, 10 µM and 100 µM. Both samples were incubated with 0.033 eq. Connectase and 0.066 eq. BcPAP for 4h at 22°C and then used for LC-MS analysis.

The samples stored at 4°C for up to 6 hours, before they were subjected (1.6 µg protein) to an Acquity BEH C4 column, using an UltiMate3000 UHPLC. They were eluted with a 0 - 50% H_2O / acetonitrile gradient in presence of 0.1% formic acid over 7 minutes. The eluted molecules were analyzed with a MaXis HD UHR q-TOF spectrometer. Mass spectrometer parameters were adapted to the size of the molecule and the chromatography flow rate (by default 0.15 ml/min). Data analysis was performed using Bruker Compass DataAnalysis v6.1 software (Bruker Daltonik,

Bremen, Germany). Charge deconvolution of the m/z spectra was performed with the MaxEnt deconvolution algorithm (Bruker Daltonic, Bremen, Germany). Deconvolution artifacts without m/z series were excluded.

Data vailability

All primary data are available in Dataset S2. All relevant data are also available from the corresponding author upon request.

Supplement

Figure S1

Visualization of the educt and product quantities in Figure 3A-F.

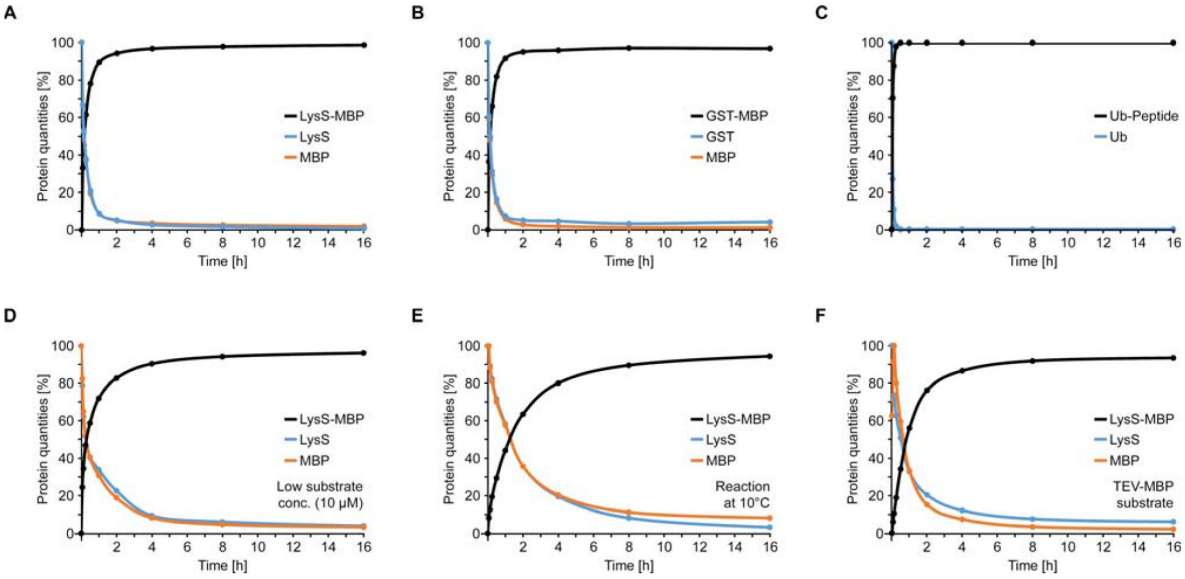


Figure S2

(related to Figure 3): Complete protein-protein ligations.

Shown are SDS-PAGE time course analyses of ligation reactions using Lysine-tRNA ligase (LysS; C-terminal ELASKDPGAFDADPLVVEI sequence) and Ubiquitin (Ub; N-terminal AGAFDADPLVVEI (A) or PAGAFDADPLVVEI (B) sequence) as substrates. Both reactions were performed with 1 eq. substrates (100 µM), 0.033 eq. Connectase, and 0.066 eq. BcPAP at 22°C. A densitometric analysis of the protein bands is shown below each experiment. For the substrates, the values reflect the substrate band density relative to the substrate band in the control sample (0 min); for the products, the values reflect the product band density relative to the total band density (substrates + products). The exact values can be found in Dataset S2.

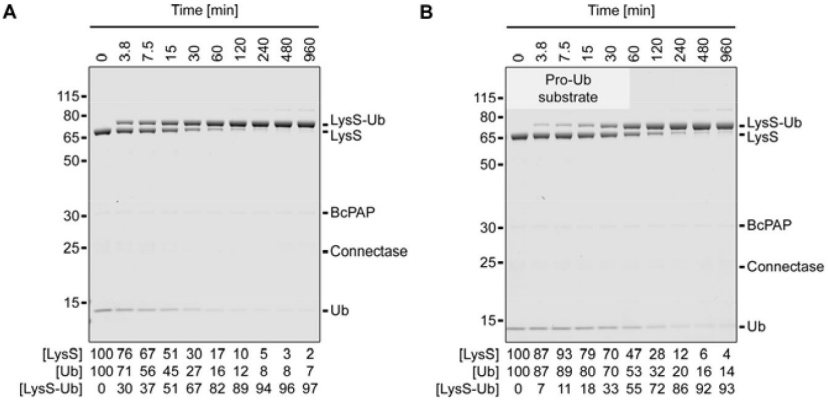


Figure S3

LC-MS analysis of an equimolar Ub-MBP mixture before (A) and after (B, C) conjugation.

A mixture of 100 µM Ub-ELASKDPGAFDADPLVVEI-Strep and 100 µM AGAFDADPLVVEI-MBP was analyzed before (A) and after (B) incubation with 0.033 eq. Connectase and 0.066 eq. BcPAP. The reaction byproduct GAFDADPLVVEI-Strep (C) appeared as an extra peak upon conjugation. The signal intensities in the plots were normalized to the most intense peak. MBP was detected both as a full-length version and as an N-terminally truncated version (MBP Δ1-221) without Connectase recognition sequence. The truncated version was not detected by SDS-PAGE (**Figure 3**), suggesting a low abundance in the sample. All detected masses can be found in Dataset S2.

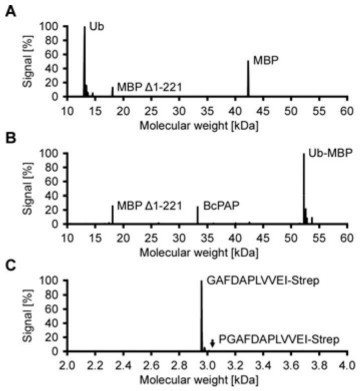


Figure S4

Effect of protease inhibitors on BcPAP activity.

Shown is the conjugation of Ub-Strep (educt, 1 eq.) to AGAFDADPLVVEI peptide (1 eq.) in presence of different protease inhibitors. The reaction catalyzed by Connectase (upper gel) is not inhibited by these substances and results in an equilibrium between Ub-Strep educt and Ub-Peptide product (as in **Figure 2**). In a reaction with Connectase and BcPAP (lower gel), up to 100% Ub-peptide product is formed. Lower product yields indicate an inhibition of BcPAP. This effect is most pronounced for the serine protease inhibitors AEBSF, PMSF, and a commercial AEBSF-containing inhibitor mix ("complete"). $ZnCl_2$, which had been reported previously as a BcPAP inhibitor[24], led to the precipitation of Connectase and BcPAP.

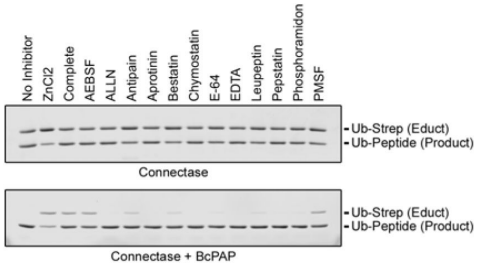


Figure S5

Quantification of αHER2 antibodies in cell culture medium.

Heavy (HC) and light (LC) antibody chains with a C-terminal Connectase recognition sequence were expressed in HEK293 cells. The medium with the exported antibodies was exchanged daily, allowing the monitoring of antibody expression levels by in-gel fluorescence (A). In this western blot alternative, Connectase is used to fuse fluorophores to the target proteins (HC, LC) and a reference protein (Ref). By comparing the intensity of the resulting fluorescent bands, daily antibody expression levels can be estimated (lower panel, see methods). A Coomassie stain of the same gel (B) shows all proteins in the cell culture medium samples. The experiment is described and discussed in detail in a previously published paper [31]. It is also depicted here because it shows the production of antibodies used in **Figure 4** and highlights the use of the Connectase recognition sequence on a protein of interest for different applications: protein detection and quantification (this figure), and protein conjugation (**Figure 4**).

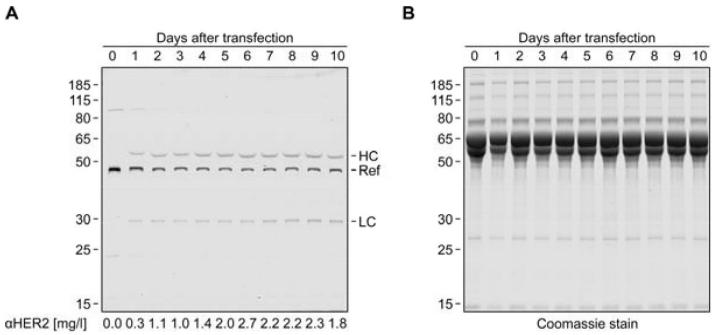


Figure S6

Time course of an antibody biotinylation reaction (related to Figure 4).

The heavy chain (HC) and light chain (LC) of αHER2 (human epidermal growth factor receptor 2) antibodies were produced with a C-terminal Connectase recognition sequence and a Streptavidin tag (HC-Strep, LC-Strep). In the reactions (25 µM αHER2, 100 µM subunits), 0.033 eq. Connectase, 0.066 eq. BcPAP, 22°C), the Streptavidin tag is replaced by a shorter biotinylated peptide (1 eq.). The figure shows an SDS-PAGE analysis of the reaction time course. Below the gel, a densitometric quantification of product bands relative to educt bands is presented. For this calculation, the BcPAP density from the control lane (Cntrl) was subtracted from the combined LC-Strep/BcPAP band.

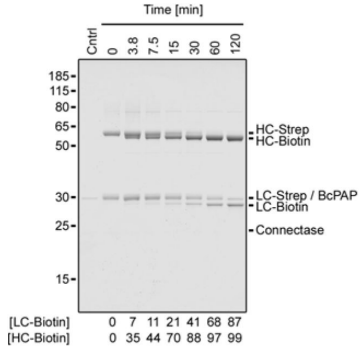


Figure S7

Stability of antibody conjugates in human plasma.

Antibodies containing C-terminal Connectase recognition sequences were biotinylated (1 µM HC_2LC_2-$Biotin_4$) and mixed with human plasma (containing ~86 µM unconjugated HC_2LC_2). The mixture was incubated for up to 7 days at 4°C, 22°C, or 37°C, then centrifuged, and the supernatant analyzed by SDS-PAGE.
(A)Coomassie G-250 staining provides a nonspecific detection of all proteins.
(B)In-gel fluorescence[17,32] specifically detects proteins with Connectase recognition sequences, i.e., HC-Biotin and LC-Biotin, but shows no bands for the plasma control sample without HC_2LC_2-$Biotin_4$ (first lane). Throughout the time course, both the HC and LC bands remain visible with consistent intensity, indicating that the biotinylated Connectase recognition sequence (ELASKDAGAFDADPLVVEIK-Biotin) remains intact under all tested conditions.

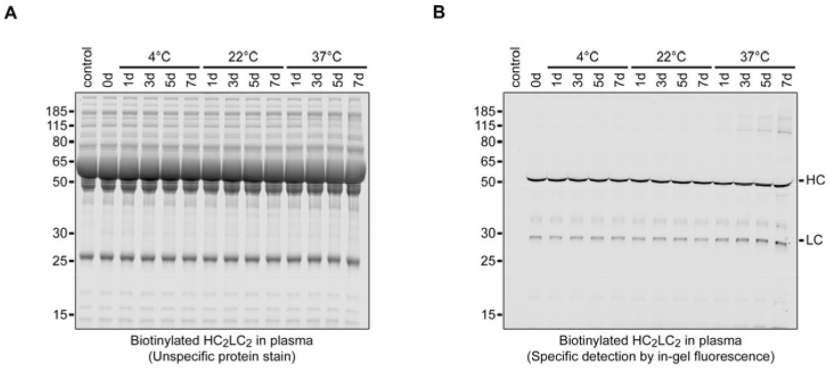
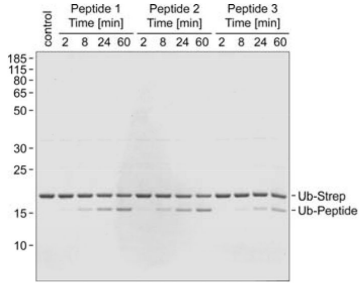


Figure S8

Determination of the minimal Connectase recognition sequence.

Connectase acts on a linker sequence derived from its physiological interaction partner, Methyltransferase A (MtrA). In an initial characterization[16], this sequence was identified as RELASKDPGAFDADPLVVEI. It remained unclear, whether it could be further shortened from the N-terminal side. The depicted gel shows the Connectase-mediated conjugation of Ub-Strep to RELASKDPGAFDADPLVVEI (peptide 1), ELASKDPGFDADPLVVEI (peptide 2), or LASKDPGAFDADPLVVEI (peptide 3). The product, Ub-peptide, is formed at a similar rate with peptide 1 (relative ligation rate determined by densitometric analysis: 94%) and peptide 2 (100%), but at a reduced rate when using peptide 3 (47%). This suggests that ELASKDPGAFDADPLVVEI is sufficient for efficient conjugation reactions. The protein substrates employed in this paper have the C-terminal RELASKDPGAFDADPLVVEI sequence, with the additional N-terminal arginine serving as a small linker.



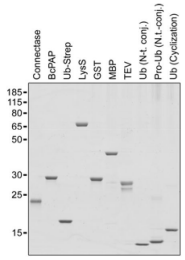


Figure S9

Proteins used in this study.

Acknowledgements

Liquid chromatography-mass spectrometry was performed by Sandra Maier and Dr. Anne Zeck at the NMI Natural and Medical Sciences Institute at the University of Tübingen, using instrumentation acquired through the Baden-Württemberg Ministry of Economic Affairs, Labor and Tourism (Germany) grant program “Special investment program for climate-neutral business-related research” (WM3-4332-3/6). We thank Andrei Lupas for discussions and continuous support. We thank Valeria Hatskovska for support in handling the eukaryotic cell cultures. This work was supported by institutional funds from the Max Planck Society and by the German Research Foundation (DFG project number 512378754 to A.C.D.F.).

Additional files

Supplementary Datasets

References

1 Stephanopoulos N., Francis M. B. 2011) **Choosing an effective protein bioconjugation strategy** *Nat Chem Biol* **7**:876–884 https://doi.org/10.1038/nchembio.720

2 Kaur J., Saxena M., Rishi N. 2021) **An Overview of Recent Advances in Biomedical Applications of Click Chemistry** *Bioconjug Chem* **32**:1455–1471 https://doi.org/10.1021/acs.bioconjchem.1c00247

3 Kolb H. C., Finn M. G., Sharpless K. B. 2001) **Click Chemistry: Diverse Chemical Function from a Few Good Reactions** *Angew Chem Int Ed Engl* **40**:2004–2021 https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5

4 Zakeri B., et al. 2012) **Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin** *Proc Natl Acad Sci U S A* **109**:E690–697 https://doi.org/10.1073/pnas.1115485109

5 Sutherland A. R., Alam M. K., Geyer C. R. 2019) **Post-translational Assembly of Protein Parts into Complex Devices by Using SpyTag/SpyCatcher Protein Ligase** *Chembiochem* **20**:319–328 https://doi.org/10.1002/cbic.201800538

6 Wu H., Hu Z., Liu X. Q. 1998) **Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp** *PCC6803. Proc Natl Acad Sci U S A* **95**:9226–9231 https://doi.org/10.1073/pnas.95.16.9226

7 Anastassov S., Filo M., Khammash M. 2024) **Inteins: A Swiss army knife for synthetic biology** *Biotechnol Adv* **73**:108349 https://doi.org/10.1016/j.biotechadv.2024.108349

8 Schmidt M., Toplak A., Quaedflieg P. J., Nuijens T. 2017) **Enzyme-mediated ligation technologies for peptides and proteins** *Curr Opin Chem Biol* **38**:1–7 https://doi.org/10.1016/j.cbpa.2017.01.017

9 Morgan H. E., Turnbull W. B., Webb M. E. 2022) **Challenges in the use of sortase and other peptide ligases for site-specific protein modification** *Chem Soc Rev* **51**:4121–4145 https://doi.org/10.1039/d0cs01148g

10 Mazmanian S. K., Liu G., Ton-That H., Schneewind O. 1999) **Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall** *Science* **285**:760–763 https://doi.org/10.1126/science.285.5428.760

11 Heck T., Pham P. H., Yerlikaya A., Thony-Meyer L., Richter M. 2014) **Sortase A catalyzed reaction pathways: a comparative study with six SrtA variants** *Catal Sci Technol* **4**:2946–2956 https://doi.org/10.1039/c4cy00347k

12 Frankel B. A., Kruger R. G., Robinson D. E., Kelleher N. L., McCafferty D. G. 2005) **Staphylococcus aureus sortase transpeptidase SrtA: Insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism** *Biochemistry* **44**:11188–11200 https://doi.org/10.1021/bi050141j

13 Jacobitz A. W., Kattke M. D., Wereszczynski J., Clubb R. T. 2017) **Sortase Transpeptidases: Structural Biology and Catalytic Mechanism** *Adv Protein Chem Str* **109**:223–264 https://doi.org/10.1016/bs.apcsb.2017.04.008

14 Chen I., Dorr B. M., Liu D. R. 2011) **A general strategy for the evolution of bond-forming enzymes using yeast display** *Proc Natl Acad Sci U S A* **108**:11399–11404 https://doi.org/10.1073/pnas.1101046108

15 Chen L., et al. 2016) **Improved variants of SrtA for site-specific conjugation on antibodies and proteins with high efficiency** *Sci Rep* **6**:31899 https://doi.org/10.1038/srep31899

16 Fuchs A. C. D., et al. 2021) **Archaeal Connectase is a specific and efficient protein ligase related to proteasome β subunits** *Proceedings of the National Academy of Sciences* **118**:e2017871118 https://doi.org/10.1073/pnas.2017871118

17 Fuchs A. C. D. 2023) **Specific, sensitive and quantitative protein detection by in-gel fluorescence** *Nat Commun* **14**:2505 https://doi.org/10.1038/s41467-023-38147-8

18 Lapteva Y. S., et al. 2021) **In Vitro N-Terminal Acetylation of Bacterially Expressed Parvalbumins by N-Terminal Acetyltransferases from Escherichia coli** *Appl Biochem Biotechnol* **193**:1365–1378 https://doi.org/10.1007/s12010-020-03324-8

19 Bandyopadhyay A., Cambray S., Gao J. 2016) **Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation** *Chem Sci* **7**:4589–4593 https://doi.org/10.1039/C6SC00172F

20 Gonzales T., Robert-Baudouy J. 1996) **Bacterial aminopeptidases: properties and functions** *FEMS Microbiol Rev* **18**:319–344 https://doi.org/10.1111/j.1574-6976.1996.tb00247.x

21 Cunningham D. F., O'Connor B. 1997) **Proline specific peptidases** *Biochim Biophys Acta* **1343**:160–186 https://doi.org/10.1016/s0167-4838(97)00134-9

22 Dong Z. X., et al. 2022) **Prolyl aminopeptidases: Reclassification, properties, production and industrial applications** *Process Biochem* **118**:121–132 https://doi.org/10.1016/j.procbio.2022.04.025

23 Yoshimoto T., Tsuru D. 1985) **Proline iminopeptidase from Bacillus coagulans: purification and enzymatic properties** *J Biochem* **97**:1477–1485 https://doi.org/10.1093/oxfordjournals.jbchem.a135202

24 Kitazono A., Yoshimoto T., Tsuru D. 1992) **Cloning, sequencing, and high expression of the proline iminopeptidase gene from Bacillus coagulans** *J Bacteriol* **174**:7919–7925 https://doi.org/10.1128/jb.174.24.7919-7925.1992

25 Kitazono A., Ito K., Yoshimoto T. 1994) **Prolyl aminopeptidase is not a sulfhydryl enzyme: identification of the active serine residue by site-directed mutagenesis** *J Biochem* **116**:943–945 https://doi.org/10.1093/oxfordjournals.jbchem.a124649

26 Beck A., Goetsch L., Dumontet C., Corvaia N. 2017) **Strategies and challenges for the next generation of antibody-drug conjugates** *Nat Rev Drug Discov* **16**:315–337 https://doi.org/10.1038/nrd.2016.268

27 Liu J., Barfield R. M., Rabuka D. 2019) **Site-Specific Bioconjugation Using SMARTag((R)) Technology: A Practical and Effective Chemoenzymatic Approach to Generate Antibody-**

Drug Conjugates *Methods Mol Biol* **2033**:131–147 https://doi.org/10.1007/978-1-4939-9654-4_10

28 Wijdeven M. A., et al. 2022) **Enzymatic glycan remodeling-metal free click (GlycoConnect) provides homogenous antibody-drug conjugates with improved stability and therapeutic index without sequence engineering** *MAbs* **14**:2078466 https://doi.org/10.1080/19420862.2022.2078466

29 Sadowsky J. D., et al. 2017) **Development of Efficient Chemistry to Generate Site-Specific Disulfide-Linked Protein- and Peptide-Payload Conjugates: Application to THIOMAB Antibody-Drug Conjugates** *Bioconjug Chem* **28**:2086–2098 https://doi.org/10.1021/acs.bioconjchem.7b00258

30 Zimmerman E. S., et al. 2014) **Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system** *Bioconjug Chem* **25**:351–361 https://doi.org/10.1021/bc400490z

31 Gebleux R., Briendl M., Grawunder U., Beerli R. R. 2019) **Sortase A Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody-Drug Conjugates** *Methods Mol Biol* **2012**:1–13 https://doi.org/10.1007/978-1-4939-9546-2_1

32 Fuchs A. C. D. 2024) **Detection and quantification of C-terminally tagged proteins by in-gel fluorescence** *Sci Rep* **14**:15697 https://doi.org/10.1038/s41598-024-66132-8

33 Gramlich M., et al. 2021) **HDX-MS for Epitope Characterization of a Therapeutic ANTIBODY Candidate on the Calcium-Binding Protein Annexin-A1** *Antibodies (Basel)* **10** https://doi.org/10.3390/antib10010011

34 Purkayastha A., Kang T. J. 2019) **Stabilization of Proteins by Covalent Cyclization** *Biotechnol Bioproc E* **24**:702–712 https://doi.org/10.1007/s12257-019-0363-4

35 Fierer J. O., Veggiani G., Howarth M. 2014) **SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture** *Proc Natl Acad Sci U S A* **111**:E1176–1181 https://doi.org/10.1073/pnas.1315776111

36 van 't Hof W., Hansenova Manaskova S., Veerman E. C., Bolscher J. G. 2015) **Sortase-mediated backbone cyclization of proteins and peptides** *Biol Chem* **396**:283–293 https://doi.org/10.1515/hsz-2014-0260

37 Arnott Z. L. P., et al. 2024) **Quantitative N-or C-Terminal Labelling of Proteins with Unactivated Peptides by Use of Sortases and a d-Aminopeptidase** *Angew Chem Int Ed Engl* **63**:e202310862 https://doi.org/10.1002/anie.202310862

38 Antos J. M., et al. 2017) **Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation** *Curr Protoc Protein Sci* **89**:15–15 https://doi.org/10.1002/cpps.38

39 Gramlich M., et al. 2022) **A Novel PNGase Rc for Improved Protein N-Deglycosylation in Bioanalytics and Hydrogen-Deuterium Exchange Coupled With Mass Spectrometry Epitope Mapping under Challenging Conditions** *Anal Chem* **94**:9863–9871 https://doi.org/10.1021/acs.analchem.2c01748

Author information

Adrian CD Fuchs
Department of Protein Evolution, Max Planck Institute for Biology, Tübingen, Germany
ORCID iD: 0000-0001-6550-1795

For correspondence: adrian.fuchs@tuebingen.mpg.de

Editors

Reviewing Editor
Yamini Dalal
National Cancer Institute, Bethesda, United States of America

Senior Editor
Yamini Dalal
National Cancer Institute, Bethesda, United States of America

Reviewer #1 (Public review):

Fuchs describes a novel method of enzymatic protein-protein conjugation using the enzyme Connectase. The author is able to make this process irreversible by screening different Connectase recognition sites to find an alternative sequence that is also accepted by the enzyme. They are then able to selectively render the byproduct of the reaction inactive, preventing the reverse reaction, and add the desired conjugate with the alternative recognition sequence to achieve near-complete conversion. I agree with the authors that this novel enzymatic protein fusion method has several applications in the field of bioconjugation, ranging from biophysical assay conduction to therapeutic development. Previously the author has published on the discovery of the Connectase enzymes and has shown its utility in tagging proteins and detecting them by in-gel fluorescence. They now extend their work to include the application of Connectase in creating protein-protein fusions, antibody-protein conjugates, and cyclic/polymerized proteins. As mentioned by the author, enzymatic protein conjugation methods can provide several benefits over other non-specific and click chemistry labeling methods. Connectase specifically can provide some benefits over the more widely used Sortase, depending on the nature of the species that is desired to be conjugated. Overall, this method provides a novel, reproducible way to enzymatically create protein-protein conjugates.

The manuscript is well-written and will be of interest to those who are specifically working on chemical protein modifications and bioconjugation.

Comments on revisions:

The authors have improved the manuscript significantly by clarifying the questions raised adding new text, providing additional references and/or adding additional data. The thorough study and efficiency of the method for enzymatic protein-protein conjugation using the enzyme Connectase warrants publication of this manuscript in its current form.

https://doi.org/10.7554/eLife.102765.2.sa2

Reviewer #2 (Public review):

Summary:

Unlike previous traditional protein fusion protocols, the author claims their proposed new method is fast, simple, specific, reversible, and results in a complete 1:1 fusion. A multi-disciplinary approach from cloning and purification, biochemical analyses, and proteomic mass spec confirmation revealed fusion products were achieved.

Strengths:

The author provides convincing evidence that an alternative to traditional protein fusion synthesis is more efficient with 100% yields using connectase. The author optimized the protocol's efficiency with assays replacing a single amino acid and identification of a proline aminopeptidase, Bacilius coagulans (BcPAP), as a usable enzyme to use in the fusion reaction. Multiple examples including Ubiquitin, GST, and antibody fusion/conjugations reveal how this method can be applied to a diverse range of biological processes.

Weaknesses:

Though the ~100% ligation efficiency is an advancement, the long recognition linker may be the biggest drawback. For large native proteins that are challenging/cannot be synthesized and require multiple connectase ligation reactions to yield a complete continuous product, the multiple interruptions with long linkers will likely interfere with protein folding, resulting in non-native protein structures. This method will be a good alternative to traditional approaches as the author mentioned but limited to generating epitope/peptide/protein tagged proteins, and not for synthetic protein biology aimed at examining native/endogenous protein function in vitro.

https://doi.org/10.7554/eLife.102765.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Reviewer #1 (Public review):***
>
> *Fuchs describes a novel method of enzymatic protein-protein conjugation using the enzyme Connectase. The author is able to make this process irreversible by screening different Connectase recognition sites to find an alternative sequence that is also accepted by the enzyme. They are then able to selectively render the byproduct of the reaction inactive, preventing the reverse reaction, and add the desired conjugate with the alternative recognition sequence to achieve near-complete conversion. I agree with the authors that this novel enzymatic protein fusion method has several applications in the field of bioconjugation, ranging from biophysical assay conduction to therapeutic development. Previously the author has published on the discovery of the Connectase enzymes and has shown its utility in tagging proteins and detecting them by in-gel fluorescence. They now extend their work to include the application of Connectase in creating protein-protein fusions, antibody-protein conjugates, and cyclic/polymerized proteins. As mentioned by the author, enzymatic protein conjugation methods can provide several benefits over other non-specific and click chemistry labeling methods. Connectase specifically can provide some benefits over the more widely used Sortase, depending on the nature of the species that is desired to be conjugated. However, due to a similar lengthy sequence between conjugation partners, the method described in this paper does not provide clear benefits over the existing SpyTag-SpyCatcher conjugation*

> *system. Additionally, specific disadvantages of the method described are not thoroughly investigated, such as difficulty in purifying and separating the desired product from the multiple proteins used. Overall, this method provides a novel, reproducible way to enzymatically create protein-protein conjugates.*
>
> *The manuscript is well-written and will be of interest to those who are specifically working on chemical protein modifications and bioconjugation.*

I'd like to comment on two points.

(1) The benefits over the SpyTag-SpyCatcher system. Here, the conjugation partners are fused via the 12.3 kDa SpyCatcher protein, which is considerably larger than the Connectase fusion sequence (19 aa). This is mentioned in the introduction (p. 1 ln 24-26). Furthermore, SpyTag-SpyCatcher fusions are truly irreversible, while Connectase/BcPAP fusions may be reversed (p. 8, ln 265-273). For example, target proteins (e.g., AGAFDADPLVVEI-Protein) may be covalently fused to functionalized magnetic beads (e.g., Bead-ELASKDPGAFDADPLVVEI) in order to perform a pulldown assay. After the assay, the target protein and any bound interactors could be released from the beads by the addition of a Connectase / peptide (AGAFDAPLVVEI) mixture.

In a related technology, the SpyTag-SpyCatcher system was split into three components, SpyLigase, SpyTag and KTag (Fierer et al., PNAS 2014). The resulting method introduces a sequence between the fusion partners (SpyTag (13aa) + KTag (10aa)), which is similar in length to the Connectase fusion sequence (p. 8, ln 297 - 298). Compared to the original method, however, this approach seems to require longer incubation times, while yielding less fusion product (Fierer et al., Figure 2).

(2) Purification of the fusion product. The method is actually advantageous in this respect, as described in the discussion (p. 8, ln 258-264). Examples are now provided in Figure 6.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *Unlike previous traditional protein fusion protocols, the author claims their proposed new method is fast, simple, specific, reversible, and results in a complete 1:1 fusion. A multi-disciplinary approach from cloning and purification, biochemical analyses, and proteomic mass spec confirmation revealed fusion products were achieved.*
>
> *Strengths:*
>
> *The author provides convincing evidence that an alternative to traditional protein fusion synthesis is more efficient with 100% yields using connectase. The author optimized the protocol's efficiency with assays replacing a single amino acid and identification of a proline aminopeptidase, Bacilius coagulans (BcPAP), as a usable enzyme to use in the fusion reaction. Multiple examples including Ubiquitin, GST, and antibody fusion/conjugations reveal how this method can be applied to a diverse range of biological processes.*
>
> *Weaknesses:*
>
> *Though the ~100% ligation efficiency is an advancement, the long recognition linker may be the biggest drawback. For large native proteins that are challenging/cannot be synthesized and require multiple connectase ligation reactions to yield a complete continuous product, the multiple interruptions with long linkers will likely interfere with protein folding, resulting in non-native protein structures. This method will be a good alternative to traditional approaches as the author mentioned but limited to generating*

> *epitope/peptide/protein tagged proteins, and not for synthetic protein biology aimed at examining native/endogenous protein function in vitro.*

The assessment is fair, and I have no further comments to add.

> ***Reviewer #1 (Recommendations for the authors):***
>
> *Major/Experimental Suggestions:*
>
> *(1) Throughout the paper only one reaction shown via gels had 100% conversion to desired product (Figure 3C). It is misleading to title a paper with absolutes such as "100% product yield", when the majority of reactions show >95% product yield, without any purification. Please change the title of the manuscript to something along the lines of "Novel Irreversible Enzymatic Protein Fusions with Near-Complete Product Yield".*

The conjugation reaction is thermodynamically favored. It is driven by the hydrolysis of a peptide bond (P|GADFDADPLVVEI), which typically releases 8 - 16 kJ/mol energy. This should result in a >99.99% complete reaction (DG° = -RT ln (Product/Educt)). In line with this, 99% - 100% of the less abundant educts (LysS, Figure 3A; MBP, Figure 3B; Ub-Strep, Figure 3C) are converted in the time courses (Figure 3D-F show different reaction conditions, which slow down conjugate formation). 100% conversion are also shown in Figure 5, Figure 6, and Figure S4. Likewise, 99.6% relative fusion product signal intensity in an LCMS analysis (Figure S2) after 4h reaction time (0.13% and 0.25% educts). In this experiment, the proline had been removed from 99.8% of the peptide byproducts (P|GADFDADPLVVEI). It is clear that this reaction is still ongoing and that >99.99% of the prolines will be removed from the peptides in time. These findings suggest that the conjugation reaction gradually slows down the less educt is available, but eventually reaches completion.

For some experiments, lower product yields (e.g. 97% in Figure 3B) are reported in the paper. These were calculated with Yield = 100% x Product / (Educt1 + Educt 2 + Product). With this formula, 100% conjugation can only be achieved with exactly equimolar educt quantities, because both educt 1 and educt 2 need to be converted entirely. If one educt 1 is available in excess, for example because of protein concentration measurement inaccuracies or pipetting errors, some of it will be left without fusion partner. In case of Figure 3B, 3% more GST seemed to have been in the mixture. These are methodological inaccuracies.

> *(2) Please provide at least one example of a purified desired product, and mention the difficulties involved as a disadvantage to this particular method. Separating BcPAP, Connectase, and the desired protein-protein conjugate may prove to be quite difficult, especially when Connectase cleaves off affinity tags.*

Examples are now provided in Figure 6. As described in the discussion (p. 8, ln 258-264), the simple product purification is one of the advantages of the method.

> *(3) For the antibody conjugate, please provide an example of conjugating an edduct that would prove to be more useful in the context of antibodies. For example, as you mention in the introduction, conjugation of fluorophores, immobilization tags such as biotin, and small molecule linker/drugs are useful bioconjugates to antibodies.*

Antibody-biotinylation is now shown in Figure S6; Antibody-fluorophore conjugates are part of Figures S5 and S7.

> *(4) Please assess the stability of these protein-protein conjugates under various conditions (temperature, pH, time) to ensure that the ligation via Connectase is stable over a broad array of conditions. In particular, a relevant antibody-conjugate stability*

> *assay should be done over the period of 1-week in both buffer and plasma to show applicability for potential therapeutics.*

The stability of an antibody-biotin conjugate in blood plasma over 7 days at different temperatures is now shown in Figure S7.

Generally, Connectase introduces a regular peptide bond (Asp-Ala) with a high chemical and physical stability (e.g. 10 min incubation at 95°C in SDS-PAGE loading buffer; H2O-formic acid / acetonitrile gradients for LC-MS). The sequence may be susceptible to proteases, although this is not the case in HEK293 cells (antibody expression), *E. coli*, or blood plasma (Figure S7).

> *(5) Please conduct functional assays with the antibody-protein/peptide conjugates to show that the antibody retains binding capabilities to the HER-2 antigen and the modification was site-selective, not interfering with the binding paratope or binding ability of the antibody in any way. This can be done through bio-layer interferometry, surface plasmon resonance, ELISA, etc.*

We plan the immobilization of the HER2 antibody on microplates and its use in an ELISA. However, this experiment requires significant testing and optimizations. It will be part of a future paper on the use of Connectase for protein immobilization.

For now, the mass spectrometry data provide clear evidence of a single site-selective conjugation, as the C-terminal ELASKDPGAFDADPLVVEI-Strep sequence is replaced by ELASKDAGAFDADPLVVEI(-Ub). Given that the conjugation sites at the C-termini are far from the antigen binding sites, and have already been used in a number of other approaches (e.g., SpyTag, SnapTag, Sortase), it appears unlikely that these conjugations interfere with antigen binding.

> *(6) Please include gels of all proteins used in ligation reactions after purification steps in the SI to show that each species was pure.*

The pure proteins are now shown in Figure S9.

> *(7) Please provide the figures (not just tables) of LC/MS deconvoluted mass spectra graphs for all conjugates, either in the main text or the SI.*

Please specify which spectra you are missing. I believe all relevant spectra are shown in Figures 4, 5, and S3. The primary data can be found in Dataset S2.

> *(8) Please provide more information in the methods section on exactly how the densitometry quantification of gel bands was performed with ImageJ.*

Details on the quantification with Image Studio Lite 5.2 were added in the method section (p. 17, ln 461-463).

> *Minor Suggestions:*
>
> *(1) Page 1, line 19: can include one sentence on what assays these particular bioconjugations are usefule for (e.g. internalization cell studies, binding assays, etc.)*

I prefer not to provide additional details here to keep the text concise and focused.

> *(2) Page 1, line 22: "three to ten equivalents" instead of 3x-10x.*

Done.

> *(3) Page 1, line 23: While NHS labeling is widely considered non-specific, maleimide conjugation to free cysteines is generally considered specific for engineered free cysteine residues, since native proteins often do not have free cysteine residues available for conjugation. If you are referring to the potential of maleimides to label lysines as well, that should be specifically stated.*

I modified the sentence, now stating that these methods are "can be" unspecific.

As pointed out, it is possible to achieve specificity by eliminating all other free cysteines and/or engineering a cysteine in an appropriate position. In many other cases, however (e.g., natural antibodies), several cysteines are available, or the sample contains other proteins/peptides. I did not want to go into more detail here and refer to the cited review.

> *(4) Page 1, line 31: "and an oligoglycine G(1-5)-B"*

Done.

> *(5) Page 1, line 34: It is not clear where in the source these specific Km values are coming from, considering these are variable based on specific conditions/substrates and tend to be reaction-specific.*

I cited another review, which lists the same values, along with a few other measurements (Jacobitz et al., Adv Protein Chem Struct Biol 2017, Table 2). It is clear that each of these measurements differs somewhat, but they are generally comparable (K_M(LPETG) = 5500 - 8760 µM; K_M(GGGGG) = 140 - 196 µM). I chose the cited study (Frankel et al., Biochemistry 2005), because it also investigated hydrolysis rates. In this study, the measurements are derived from the plots in Figure 2.

> *(6) Page 1, line 47: the comparison to western blots feels a little like apples to oranges, even though this comparison was made in previous literature. Engineering an expressed protein to have this tag and then using the tag to detect and quantify it, feels more akin to a tagging/pull down assay than a western blot in which unmodified proteins are easily detected.*

It is akin to a frequently used type of western blots with tag-specific antiboies, e.g. Anti-His_6, -Streptavidin, -His_6, -HA ,-cMyc, -Flag. I modified the sentence to clarify this.

> *(7) Page 2, line 51: "Connectase cleaves between the first D and P amino acids in the recognition sequence, resulting in an N-terminal A-ELASKD-Connectase intermediate and a C-terminal PGAFDADPLVVEI peptide."*

I prefer the current sentence, because we assume that a bond between the aspartate and Connectase is formed before PGAFDADPLVVEI is cleaved off.

> *(8) Page 3, line 94: "Exact determination is not possible due to reversibility of the reaction", the way it is stated now sounds like it is a flaw in the methods. Also, update Figure 2 to read "Estimated relative ligation rate".*

Done.

> *(9) Page 3, lines 101-107: This is worded in a confusing way. It can either be X_1 or X_2 that is inactivated depending on if the altered amino acid is on the original protein sequence or on the desired edduct to conjugate. You first give examples of how to render other*

> *amino acids inactive, but then ultimately state that proline made inactive, so separate the two distinct possibilities a bit more clearly.*

The reaction requires the inactivation of X_1, without affecting X_2 (ln 100 - 102). This is true, no matter whether it is X_1 = A, C, S, or P that is inactivated. I added a sentence to clarify this (ln 102 – 103).

> *(10) Page 4, line 118: Give a one-sentence justification for why these proteins were chosen to work with (easy to express, stable, etc).*

Done.

> *(11) Page 5, line 167: "payload molecules".*

Done.

> *(12) Page 5, lines 170-173: Word this more clearly- "full conversion with many of these methods is difficult on antibodies due to each heavy and light chain being modified separately, resulting in only a total yield of 66% DAR4 even when 90% of each chain is conjugated."*

I rephrased the section.

> *(13) Page 8, line 290: Discuss other disadvantages of this method including difficulties purifying and in incorporating such a long sequence into proteins of interest.*

Product purification is shown in the new Figure 6. As stated above, I consider the simple purification process an advantage of the method. The genetic incorporation of the sequence into proteins is a routine process and should not make any difficulties. The disadvantages of long linker sequences between fusion partners are now discussed (p.8 – 9, ln 300-302).

> *(14) Page 10, line 341: 'The experiment is described and discussed in detail in a previously published paper.31"*

Done.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *Minor Points:*
>
> *(1) It's unclear how the author derived 100% ligation rate with X = Proline in Figure 2 when there is still residual unligated UB-Strep at 96h. Please provide an expanded explanation for those not familiar with the protocol. Is the assumption made that there will be no UB-Strep if the assay was carried out beyond 96h?*

I clarified the figure legend. The assay shows the formation of an equilibrium between educts and products. Therefore, only ~50% Ub-Strep is used with X = Proline (see p. 2, ln 79 - 81). The "relative ligation rate" refers to the relative speed with which this equilibrium is established. The highest rate is seen with X = Proline, and it is set to 100%. The other rates are given relative to the product formation with X = Proline.

> *(2) Though the qualitative depiction of the data in Figure 3 is appreciated, an accompanying graphical representation of the data in the same figure will greatly enhance reception and better comprehension of several of the author's conclusions.*

Graphs are now shown in Figure S1.

> *(3) Figure 3 panel E is misaligned. Please align it with panel B above it.*

Done, thank you.

> *(4) The author refers to 'The resulting circular assemblies (37% UB2...)' in the text but identifies it as UB-C2 in Figure 5B. Is this a mistake or does UB2 refer to another assembly not mentioned in the Figures? Please check for inconsistencies.*

All circular assemblies are now labeled Ub-C $_{1\text{-}6}$.

> *(5) Finishing with a graphical schematic that depicts the entire protocol in a simple image would be much appreciated and well-received by readers. Including the scheme with A and B proteins, the recognition linkers, the addition of connectase and BcPAP, etc. to the final resulting protein with connected linker.*

A graphical summary of the reaction is now included in Figure 6.

https://doi.org/10.7554/eLife.102765.2.sa0