Archaeal Connectase is a specific and efficient protein ligase related to proteasome β subunits

Adrian C. D. Fuchs[a], Moritz Ammelburg[a,1], Jörg Martin[a], Ruth A. Schmitz[b], Marcus D. Hartmann[a], and Andrei N. Lupas[a,2]

[a]Department of Protein Evolution, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; and [b]Institute for General Microbiology, Christian Albrecht University of Kiel, 24118 Kiel, Germany

Edited by Alexander Varshavsky, California Institute of Technology, Pasadena, CA, and approved January 29, 2021 (received for review August 23, 2020)

Sequence-specific protein ligations are widely used to produce customized proteins "on demand." Such chimeric, immobilized, fluorophore-conjugated or segmentally labeled proteins are generated using a range of chemical, (split) intein, split domain, or enzymatic methods. Where short ligation motifs and good chemoselectivity are required, ligase enzymes are often chosen, although they have a number of disadvantages, for example poor catalytic efficiency, low substrate specificity, and side reactions. Here, we describe a sequence-specific protein ligase with more favorable characteristics. This ligase, Connectase, is a monomeric homolog of 20S proteasome subunits in methanogenic archaea. In pulldown experiments with *Methanosarcina mazei* cell extract, we identify a physiological substrate in methyltransferase A (MtrA), a key enzyme of archaeal methanogenesis. Using microscale thermophoresis and X-ray crystallography, we show that only a short sequence of about 20 residues derived from MtrA and containing a highly conserved KDPGA motif is required for this high-affinity interaction. Finally, in quantitative activity assays, we demonstrate that this recognition tag can be repurposed to allow the ligation of two unrelated proteins. Connectase catalyzes such ligations at substantially higher rates, with higher yields, but without detectable side reactions when compared with a reference enzyme. It thus presents an attractive tool for the development of new methods, for example in the preparation of selectively labeled proteins for NMR, the covalent and geometrically defined attachment of proteins on surfaces for cryo-electron microscopy, or the generation of multispecific antibodies.

protein ligation | sortase | transpeptidase | proteasome | methanogenic archaea

Proteins are key elements throughout biological systems, orchestrating all aspects of their function. Correspondingly, controlled protein turnover is an essential activity of all living cells, for which they have evolved a huge diversity of proteolytic enzymes, intra- and extracellularly, constitutive and inducible, sequence-specific and promiscuous, energy-dependent and -independent. One particularly prominent class among these is formed by self-compartmentalizing proteases, large multimeric complexes with barrel-shaped architectures, which segregate the proteolytic active sites to an inner compartment and control access to it with rings of nucleotidases. The prototypical representative of this class is the proteasome, central to controlled protein turnover across all branches of archaea and eukaryotes, and increasingly discovered in many lineages of bacteria (1–5). Its subunits are part of the NTN (N-terminal nucleophile) hydrolase family and form an αββα-fold, in which the active-site Thr-1 uses two nucleophilic groups—hydroxyl side chain and N-terminal amino group—for catalysis. The proteolytic subunits are usually transcribed with a propeptide, which is removed in an autocatalytic process (autolysis). Here, the Thr-1 hydroxyl group attacks the carbonyl carbon of the preceding propeptide residue, resulting in ester-bond formation (6). In a second step, the Thr-1 amino group deprotonates a water molecule for hydrolysis of that ester and, hence, activation of the enzyme. Mature proteasome subunits can then catalyze protein hydrolysis in a similar reaction, except that the Thr-1 hydroxyl group attacks the substrate carbonyl carbon instead of the propeptide carbonyl carbon.

Most other proteases share this basic mechanism but utilize separate residues to catalyze the reaction (7). For example, trypsin performs the nucleophilic attack on the substrate carbonyl carbon with the hydroxyl group of a serine and the activation of the water molecule for hydrolysis with a histidine. Many of these proteases are also flexible in terms of the employed small-molecule nucleophile: Instead of H_2O (hydrolysis), the proteasome and trypsin can also utilize the NH_2 group of a peptide N terminus (aminolysis), resulting in the ligation of two peptides (8–10). In contrast to hydrolysis, such reactions are reversible as the amount of peptide bonds remains the same: The substrate A–B is not cleaved into A and B but recombined with C to form A–C and B. Thus, this reaction could be used to connect two proteins of interest (9). Such applications, however, are difficult because water is much more abundant as a nucleophile than the amino group of substrate C and a reaction with water results in irreversible proteolysis. Moreover, most proteases are not specific enough to react with just one site in A, B, and C and would therefore generate a mixture of products. Consequently, optimized proteases, such as trypsiligase (10) or subtiligase (11), currently find use in only a narrow range of ligase applications (12).

Significance

Connectase is an unusual monomeric homolog of proteasome subunits which, unlike all other currently known members of this hydrolase family, is a sequence-specific protein ligase. We identify a natural target for it in a key enzyme of the pathway by which carbon dioxide and hydrogen are converted to methane in archaea. The transpeptidase reaction it catalyzes is highly sequence-specific, allowing it to efficiently ligate proteins that carry the recognition sequence, even in dilute and impure solutions. Connectase ligates substrates at substantially higher rates than other transpeptidases and without detectable side reactions, thus representing a valuable new tool for bioengineering.

Author contributions: M.A. discovered the homology of DUF2121 to proteasome subunits and initiated the project; R.A.S. provided the *Methanosarcina mazei* extract; A.C.D.F. designed, performed and analyzed most experiments reported here and discovered the ligase activity of DUF2121; M.D.H. solved the crystal structures; M.D.H., J.M., and A.N.L. supervised the project and discussed all results; and A.C.D.F. wrote the paper with input from all authors.

Competing interest statement: Max Planck Innovation has filed a provisional patent on Connectase and its use for enzymatic ligation.

This article is a PNAS Direct Submission.

[1]Present address: Fish & Richardson P.C., 80807 München, Germany.

[2]To whom correspondence may be addressed. Email: andrei.lupas@tuebingen.mpg.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2017871118/-/DCSupplemental.

Published March 9, 2021.

Aminolysis is not just an "accidental" side reaction of proteases but is also used to catalyze physiological ligations. The most prominent example is Sortase A (SrtA), which ligates ("sorts") exported cell wall proteins (13). By virtue of having evolved as a protein ligase, this enzyme shows favorable characteristics in bioconjugations, although it still retains some proteolytic activity. It appears that this side reaction persists despite opposing evolutionary pressure (14), demonstrating how deeply related ligase and protease reactions are and how hard they are to separate.

Here, we describe a distant, monomeric proteasome homolog which connects proteins in a sequence-specific manner and which we therefore name Connectase. Unlike all other protein ligases in use (12), Connectase forms a hydrolysis-resistant amide intermediate between the Thr-1 amino group and the substrate carbonyl carbon, allowing protein ligations without side reactions. In the following, we explore the characteristics of those ligations and discuss potential applications.

Results

Connectase Is a Proteasome Homolog of Methanogenic Archaea. We study the emergence of molecular complexity in evolution and employ the proteasome as one of our model systems (15). While exploring the prokaryotic proteasome homologs Anbu and BPH (16, 17), we identified a representative of this family denoted as Domain of Unknown Function (DUF) 2121 in public databases and as Connectase in this work. Connectase is highly divergent from all other yet described members of the proteasome family and shares only about 11% sequence identity (~38% similarity) with its closest homologs, the archaeal proteasome beta subunits. Consequently, even sensitive sequence comparison methods failed to detect this relationship at the time and we could only substantiate it by HMM–HMM comparisons with HHpred, a tool we had newly developed (18, 19). Despite the divergence in sequence, Connectase is predicted to assume a proteasome-like fold and retains the putative active site Ser/Thr residue N-terminally after the presumed processing by methionine aminopeptidase (*SI Appendix*, Fig. S1). Representatives are found exclusively in archaea capable of producing methane from carbon dioxide and molecular hydrogen [hydrogenotrophic methanogenesis (20); *SI Appendix*, Fig. S2]. Among those, two Connectase variants exist: in class II methanogens (21), such as *Methanosarcina mazei*, the protein is composed of just the NTN domain, whereas in class I methanogens, such as *Methanocaldococcus jannaschii*, it features an extra small C-terminal beta-barrel domain with remote similarity to OB folds (oligosaccharide binding). Although the speciation event separating class I and II methanogens probably occurred more than 3 billion y ago (22), the mesophilic *M. mazei* and thermophilic *M. jannaschii* variants (34% sequence identity, 54% similarity) retain similar characteristics. For their experimental characterization, we succeeded in producing both as soluble proteins (>20 mg/mL) in *Escherichia coli* with high yield (>40 mg per liter of culture).

Connectase Binds and Modifies Methyltransferase Subunit A. In order to study the function of Connectase we performed pulldown experiments with *M. mazei* cell extract, using recombinant Strep-/Myc- or HA-tagged *M. mazei* Connectase as a bait. Mass spectrometric analysis identified several potential binding partners, among which only the interaction with the most abundant candidate, methyltransferase subunit A (MtrA), could be confirmed (*SI Appendix*, Fig. S3*A*). MtrA is part of the membrane-bound MtrA–MtrH complex and acts in the hydrogenotrophic methanogenesis pathway, where it couples carbon dioxide reduction to an energy-conserving sodium gradient (23). To validate this interaction, we coeluted a purified MtrA derivative lacking the C-terminal membrane anchor (MtrA Δ219–240) with His_6-tagged Connectase from a Ni-NTA (nitrilotriacetic acid) column (Fig. 1*A*). Surprisingly, this experiment also revealed the Thr-1–dependent formation of specific reaction products. Mass spectrometrical analysis identified these products as a C-terminal MtrA fragment ($MtrA^C$; Fig. 1*B*) and a conjugate between the corresponding N-terminal MtrA fragment and Connectase ($MtrA^N$-Connectase; Fig. 1*C*). This conjugate was 18 Da—that is, the mass of H_2O—lighter than the combined mass of its components, indicating that it presents a covalent protein adduct formed by condensation.

The MtrA modification site contains a highly conserved KDPGA motif present in almost all Connectase-encoding archaea (*SI Appendix*, Fig. S4). Connectase processes MtrA between the aspartate and proline residues of this sequence, resulting in $MtrA^C$ (i.e., the C-terminal MtrA fragment starting with PGA) and $MtrA^N$-Connectase (i.e., the N-terminal MtrA fragment with KD fused to Connectase Thr-1). Accordingly, the tryptic fusion peptide (D-TLVIAFIGK) that results from a conjugation of MtrA with the Connectase N terminus (TLVIAFIGK...) is abundant in each of the mass spectrometry datasets derived from pulldowns with whole-cell extracts (discussed above). Moreover, when all free amino groups of an MtrA-Connectase tryptic peptide mixture were methylated [dimethyl labeling (24)], modifications in the above fragment were found only at the newly generated aspartate N terminus and the lysine (Fig. 1*D*). By contrast, no methylation could be detected at the threonine, indicating that its amino group is engaged in a regular amide bond with the MtrA aspartate (*SI Appendix*, Fig. S3*B*).

Connectase Recombines Proteins via the MtrA (5)KDPGA(10) Recognition Sequence. We investigated the structural basis for this reaction with the $Connectase^{S1A}$ active-site mutant and MtrA from the thermophilic *M. jannaschii*, as *M. mazei* MtrA proved prone to precipitation and hence unsuitable for high-resolution structural analyses. Light-scattering experiments show the monomeric nature of both proteins and their tight interaction within the heterodimer (Fig. 2*A* and *SI Appendix*, Fig. S5 *A* and *B*). Microscale thermophoresis (MST) experiments with synthetic peptides containing the KDPGA motif plus x N-terminal and y C-terminal residues from the MtrA sequence, denoted as (x)KDPGA(y) in the following, show that the (0)KDPGA(10) sequence is sufficient

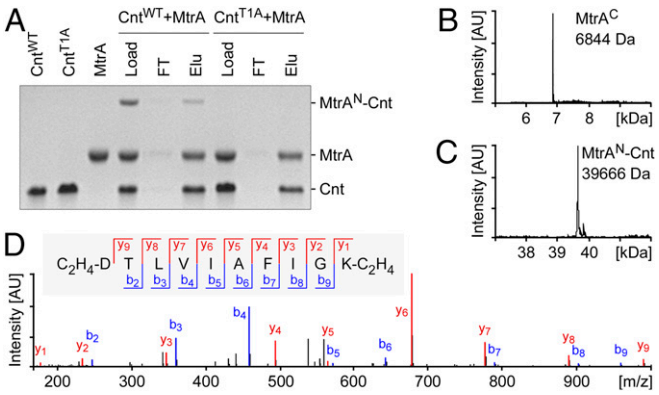


Fig. 1. Connectase (Cnt) binds and modifies MtrA. (*A*) Polyacrylamide gel showing purified $MtrA^{\Delta 219–240}$ and His_6-tagged Connectase (Cnt) proteins from *M. mazei*. When a mix of both (Load; 20 μM each) is applied on an Ni-NTA column, MtrA is not found in the flow-through (FT) but coelutes with Connectase (Elu). Furthermore, MtrA forms a specific reaction product ($MtrA^N$-Connectase; ~5 μM) with wild-type Connectase (Cnt^{WT}) but not with its active site mutant (Cnt^{T1A}). (*B* and *C*) Newly generated masses in a *M. mazei* Connectase-MtrA reaction, as identified via LC-MS. These are interpreted as $MtrA^C$ (residues 155 to 218; theoretical mass: 6,844.7 Da/detected mass: 6,843.5 Da) and as $MtrA^N$-Connectase conjugate (39,669.4 Da/39,665.8 Da). (*D*) MS/MS spectrum of the fusion peptide DTLVIAFIGK resulting from amide bond formation between the *M. mazei* Connectase N terminus (TLVIAFIGK...) and the MtrA KDPGA motif. The sample was digested with trypsin and free amino groups were dimethylated (C_2H_4). The threonine amino group is not modified, suggesting its involvement in an amide bond.

for this high-affinity interaction (Fig. 2*B* and *SI Appendix*, Fig. S5 *E–L*).

We crystallized both Connectase alone and a (15)KDPGA(10)-Connectase^{S1A} complex and could solve their structures at a resolution of 2.3 Å and 3.05 Å, respectively (Fig. 2*C* and *SI Appendix*, Fig. S5*M* and Table S1). Compared with proteasome subunits, Connectase lacks a two-helix element mediating intersubunit contacts within the proteasome complex but features a two-helix insertion at a different position, which interacts with the KDPGA sequence and might control substrate specificity. The C-terminal beta-barrel domain, which is found only in some Connectase variants, is connected via a long helical linker. It may assist in binding of MtrA (*SI Appendix*, Fig. S5 *A–D*) but is generally not required for the Connectase reaction (discussed below). Substrate binding appears to induce no major structural changes in Connectase (*SI Appendix*, Fig. S5*M*). In the NTN domain, it is mediated via induced beta-sheet interactions with four residues in the KDPGA(10) sequence, locating the particularly fragile (25) Asp–Pro bond to the active site cleft. In contrast, the crystal structure did not reveal specific interations with the helix preceding the KDPGA(10) sequence.

To verify these results in vitro, we incorporated the *M. mazei* MtrA-derived (5)KDPGA(10) sequence in a different protein, *Caldiarchaeum subterraneum* ubiquitin (26), and found that Connectase modifies it just like MtrA. Surprisingly, though, when we followed the time course of the reaction we found only a constant fraction (~7%) of Ub-(5)KDPGA(10) modified at any time (Fig. 3*B*, lanes 1 to 4). This suggested a reversible reaction, resulting in an equilibrium between modified and unmodified Ub-(5)KDPGA(10) (Fig. 3*A*). In this scenario, the "forward" reaction yields Ub-(5)KD-Connectase and PGA(10), while the "reverse" reaction restores Ub-(5)KDPGA(10). To prove this assumption, we designed alternative substrates that could be used in the "reverse" reaction, PGA(10)-sdAb (single-domain antibody), PGA(10)-CyP (cyclophilin), and PGA(10)-GST (glutathione S-transferase). These substrates could be conveniently produced via expression of MPGA(10)-sdAb/CyP/GST in *E. coli*, following "automatic" methionine removal by the endogenous methionine aminopeptidase (27). Indeed, a reaction with Ub-(5)KDPGA(10) and PGA(10)-sdAb/Cyc/GST resulted in equimolar ratios of educts and Ub-(5)KDPGA(10)-sdAb/Cyc/GST products (Fig. 3*B*, lanes 5 to 10 and *SI Appendix*, Figs. S6 and S7). Hence, Connectase appears to be capable of ligating any two substrates featuring the (5)KDPGA(10) and PGA(10) sequences in an accessible manner.

Based on the reversibility of the reaction, the identification of an amide-bonded intermediate (Fig. 1*D*), and the requirement of an active-site hydroxyl group as the catalytic nucleophile (Fig. 1*A*), we propose that the Connectase catalytic mechanism is a variation of steps involved in two known proteasomal reactions, hydrolysis and autolysis (6), but differs from them by avoiding the final irreversible hydrolysis step (Fig. 3*C*). Within the resulting equilibrium, most substrates (e.g., MtrA in Fig. 1*A*) exist in the unmodified state (left side in Fig. 3*C*) and only a small fraction is bound to Connectase via an amide bond (right side in Fig. 3*C*). The hydrolysis-sensitive ester intermediate is almost not populated (*SI Appendix*, Fig. S3*B*), providing a rationale for the apparent absence of substrate hydrolysis. The underlying reaction sequence is known as ordered ping-pong mechanism, meaning that the reaction can only start with the primary R_1-(5)KDPGA(10) substrate and only proceed with the secondary PGA(10)-R_2 substrate ($R_{1/2}$ = any molecule). As both substrates utilize the same binding site, reaction rate and product yield are influenced by primary/secondary substrate ratios. In this model, the highest molar product/educt$_1$ + educt$_2$ yield can be achieved with equimolar substrate ratios, making this the setup of choice for most protein–protein ligations. Nevertheless, the use of different substrate ratios can be useful, where complete modification of one reaction

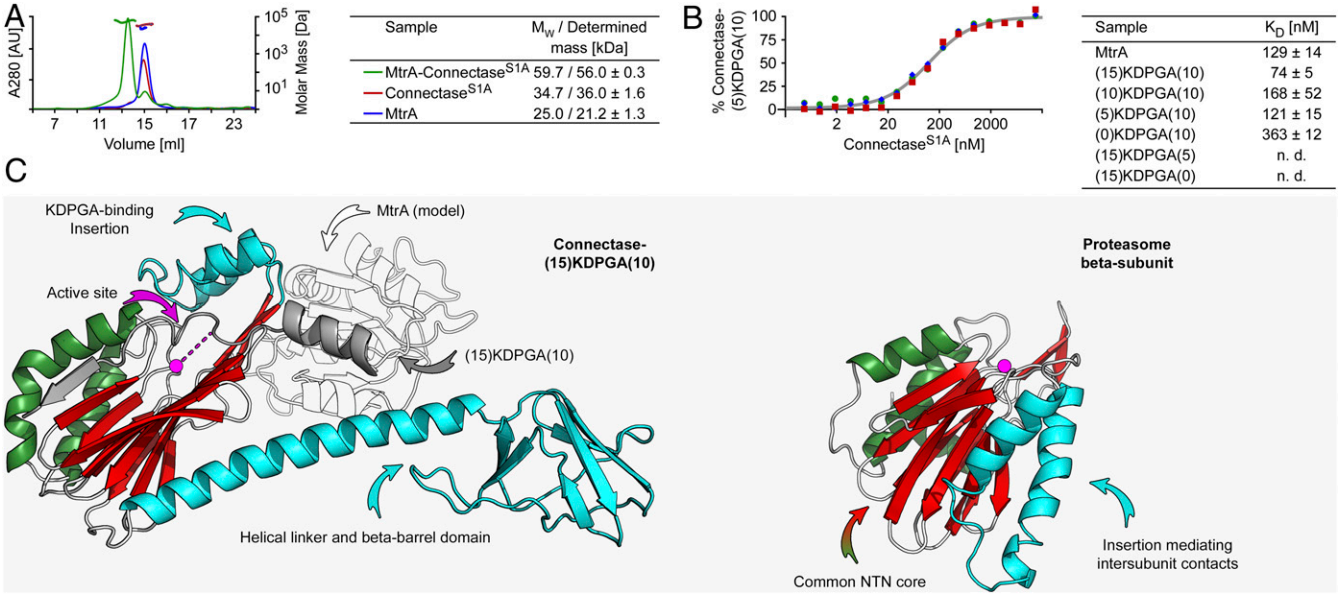


Fig. 2. MtrA interacts with Connectase via a short amino acid sequence, forming a heterodimer. (*A*) Gel filtration and light-scattering analyses of *M. jannaschii* MtrA$^{\Delta 225–245}$ and Connectase^{S1A} proteins. While Connectase and MtrA alone show a comparable elution behavior, the mixture of both elutes at a lower volume, indicating complex formation (thin lines, plotted on the primary *y* axis). This interpretation is supported by light-scattering measurements (thick lines, secondary *y* axis). The determined masses (table, *Right*) closely resemble the theoretical monomeric masses for Connectase and MtrA alone and for the MtrA–Connectase heterodimer. The values were determined in $n = 3$ independent experiments, with ± signifying the SD. (*B*) Binding curve visualizing the formation of a complex between the MtrA-derived labeled peptide (5)KDPGA(10) and *M. jannaschii* Connectase^{S1A}, as determined by MST. Analogous experiments with other substrates (table, *Right*) show that the 15-amino-acid peptide (0)KDPGA(10) is sufficient for this high-affinity interaction. The values were determined in $n = 3$ independent experiments (red, green, and blue dots), with ± signifying the SD. (*C*) Crystal structures of *M. jannaschii* Connectase^{S1A} in complex with (15)KDPGA(10) (*Left*) and the proteasome beta subunit [*Right*, PDB ID code 3H4P (57)]. Both proteins share a common NTN core domain (red and green) and an N-terminal active site residue (pink) but diverge in three protein-specific elements (cyan). A model based on an alignment with (15)KDPGA(10) shows how MtrA [colorless, PDB ID code 5L8X (23)] could potentially bind to Connectase.

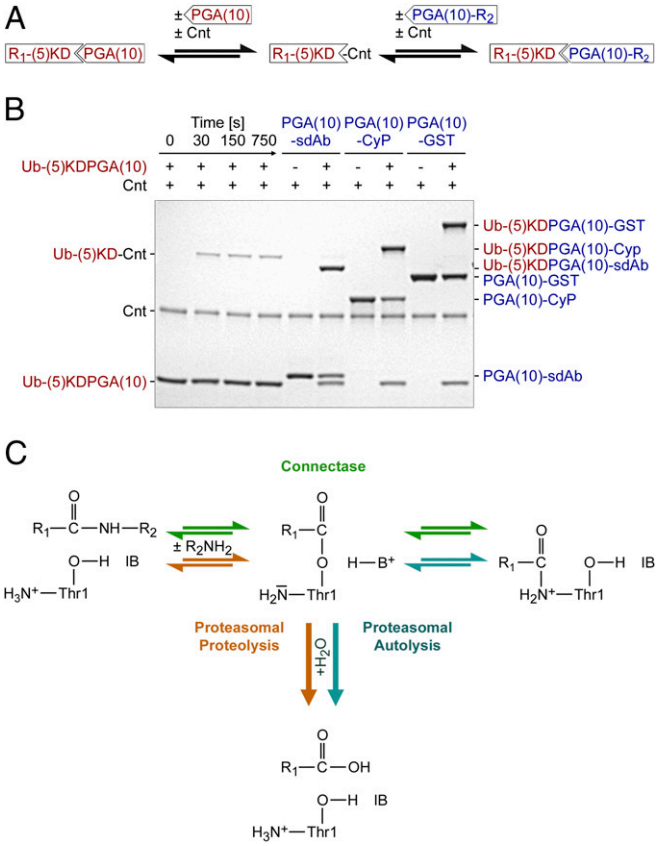


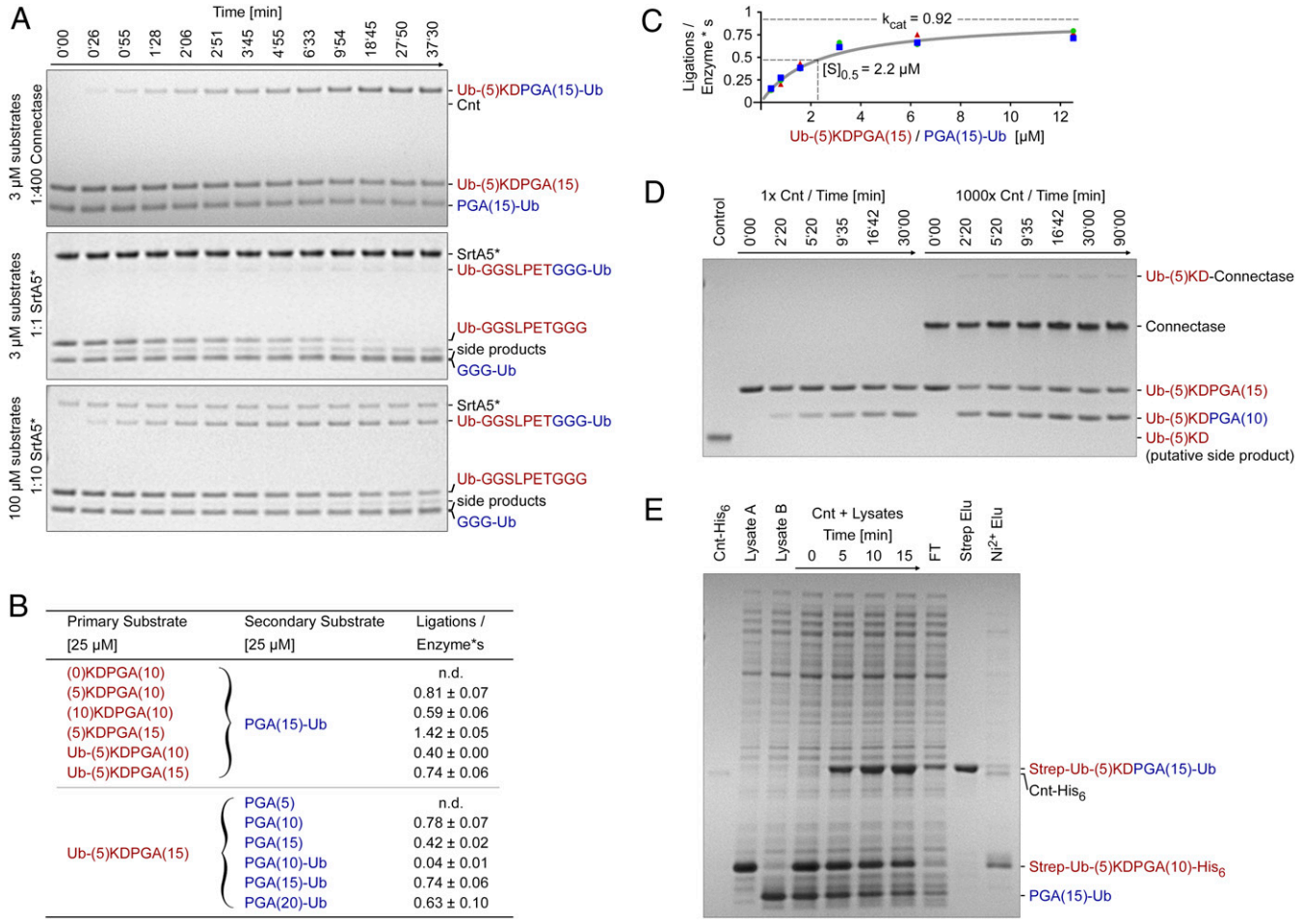
Fig. 3. Connectase (Cnt) modifications are reversible and allow the recombination of substrates. (*A*) Schematic representation of a Connectase-mediated ligation of two protein substrates, R_1-(5)KDPGA(10) and PGA(10)-R_2. (*B*) Polyacrylamide gel showing the incubation of *M. mazei* Connectase with various proteins bearing the Connectase recognition sequence. A constant fraction of Ub-(5)KDPGA(10) forms a conjugate with Connectase [Ub-(5)KD-Connectase], suggesting a reversible reaction between the two (lanes 1 to 4). In place of PGA(10), PGA(10)-sdAb/CyP/GST can be used as alternative substrates for the "reverse reaction," resulting in an equilibrium of educts and Ub-(5)KDPGA(10)-sdAb/CyP/GST products (lanes 5 to 10). A time course of these ligations in shown in *SI Appendix*, Fig. S7. (*C*) The proposed Connectase reaction mechanism (green) as a combination of the first steps of two known proteasomal reactions: proteolysis (orange) and autolysis (i.e., propeptide processing, cyan). A substrate, R_1-R_2 (*Left*), is cleaved (*Middle*) and the N-terminal fragment R_1 transferred on the Connectase Thr-1/Ser-1 N terminus (*Right*; see Fig. 1*D*). The reaction is reversible, allowing substrate recombination with an alternative R_2 substrate (*A*). It differs from proteolysis/autolysis by avoiding the irreversible hydrolysis step (*Bottom*). Some Connectase variants use Ser-1 instead of Thr-1 (shown) as active site residue (Dataset S2).

partner is desired, for example when labeling a protein with a fluorophore.

Connectase Catalyzes Specific and Efficient Ligations without Side Reactions. A more detailed characterization of the reaction shows that Connectase-catalyzed protein–protein ligations at a molar enzyme:substrate ratio of 1:400 usually result in an equilibrium of equimolar substrates and products within minutes (Fig. 4*A*). The reaction is not inhibited by classical proteasome inhibitors, most likely because they cannot mimic the highly specific Connectase recognition sequence (*SI Appendix*, Fig. S8). A systematic analysis of sequence determinants N- and C-terminally of the MtrA-derived (x)KDGPA(y)/PGA(z) motif suggests that 5 residues N-terminally of KDPGA and 10 residues C-terminally of KDPGA/PGA allow efficient ligations in most cases but that sterically demanding protein–protein ligations are much faster with 15 residues present C-terminally of PGA (Fig. 4*B* and *SI Appendix*, Fig. S9). Although these residues flanking the KDPGA motif are less well conserved (*SI Appendix*, Fig. S4), we assume that additional sequence features, which we have not determined so far, contribute to substrate recognition by Connectase.

The maximum rate of an exemplary *M. mazei* Connectase catalyzed protein–protein ligation using Ub-(5)KDPGA(15) and PGA(15)-Ub was determined with 0.92 ligations per enzyme and second (catalytic constant k_{cat}). The half-maximum reaction speed was observed at a substrate concentration of ∼2.2 μM each (Fig. 4*C* and *SI Appendix*, Fig. S10). Similar ligation rates were found in reactions with full-length *M. jannaschii* Connectase, while a deletion of its beta-barrel domain led to a slight decrease in reactivity (*SI Appendix*, Fig. S11). This finding has several implications. First, the ligase activity is conserved in two phylogenetically diverse Connectase orthologs which display maximum activity at substantially different temperatures (50 °C [*M. mazei* Connectase] and 85 °C [*M. jannaschii* Connectase]; *SI Appendix*, Fig. S12). This raises the possibility of finding useful alternatives in methanogens with lower or higher temperature optima (e.g., *Methanoculleus marisnigri*, T_{Opt} = 20 to 25 °C; *Methanopyrus kandleri*, T_{Opt} = 98 °C). Second, the small beta-barrel domain in Connectase from class I methanogens appears not to be required for the ligase reaction. Third, both a Ser-1 (in *M. jannaschii* Connectase) and a Thr-1 residue (in *M. mazei* Connectase) can serve as active-site nucleophiles (Dataset S2). In this respect, Connectase enzymes differ from other proteasome homologs studied so far, which usually use Thr-1, presumably because the Thr1-methyl group anchors its side chain in just the right orientation (6). Thus, proteasome-like domains do not have a strict requirement for Thr-1 per se but rather evolved their active sites for the use of either serine or threonine in different subfamilies.

We then sought to compare these characteristics with the most widely used enzyme ligase, SrtA, which has proven a powerful and reliable tool in numerous remarkable applications (28–31). SrtA has been extensively optimized (32, 33) and state-of-the-art in many laboratories is a SrtA pentamutant (SrtA5*) with significantly increased activity (34). Analogous to Connectase, SrtA ligates two sequences bearing the LPET-G motif and an N-terminal glycine, respectively, though additional linker sequences are usually introduced to avoid steric hindrances and to increase reactivity (14, 35, 36). In addition, SrtA also catalyzes the irreversible hydrolysis of substrates and products featuring this motif at a lower rate [$k_{cat\ Ligation}/k_{cat\ Hydrolysis} \approx 3.3$ (37)] as well as side reactions, in which lysine side chains substitute the N-terminal glycine substrate. The maximum amount of product can therefore only be obtained by monitoring the ligation/hydrolysis ratio and by stopping the reaction at just the right time. Following established protocols (36, 38), we designed Ub-GGSLPETGGG and GGG-Ub substrates and recorded the time course of the reaction. In our assays, SrtA5* displayed ∼100× higher ligase activity compared with SrtA but also catalyzed more side reactions, possibly due to its low affinity for the secondary substrate ($K_{M\ LPETG}$ = 170 μM; $K_{M\ GGG}$ = 4,700 μM (34); *SI Appendix*, Fig. S13). Unlike Connectase, both Sortase variants showed only spurious ligase activity at low substrate concentrations (3 μM each) but displayed far more favorable characteristics at high substrate concentrations (100 μM each; Fig. 4*A* and *SI Appendix*, Fig. S13). Yet, even under those conditions, Connectase shows >4,000× higher ligase activity than nonoptimized SrtA and >40× increased ligase activity compared with optimized SrtA5*. Furthermore, we observed increased ligation yields in Connectase reactions (∼50% compared with ∼30%), which we attribute to the apparent absence of side reactions. These results are comparable to Sortase ligations of other protein substrates (35, 38, 39) and demonstrate that Sortase is a



B

Primary Substrate [25 µM]	Secondary Substrate [25 µM]	Ligations / Enzyme*s
(0)KDPGA(10)	PGA(15)-Ub	n.d.
(5)KDPGA(10)	PGA(15)-Ub	0.81 ± 0.07
(10)KDPGA(10)	PGA(15)-Ub	0.59 ± 0.06
(5)KDPGA(15)	PGA(15)-Ub	1.42 ± 0.05
Ub-(5)KDPGA(10)	PGA(15)-Ub	0.40 ± 0.00
Ub-(5)KDPGA(15)	PGA(15)-Ub	0.74 ± 0.06
Ub-(5)KDPGA(15)	PGA(5)	n.d.
Ub-(5)KDPGA(15)	PGA(10)	0.78 ± 0.07
Ub-(5)KDPGA(15)	PGA(15)	0.42 ± 0.02
Ub-(5)KDPGA(15)	PGA(10)-Ub	0.04 ± 0.01
Ub-(5)KDPGA(15)	PGA(15)-Ub	0.74 ± 0.06
Ub-(5)KDPGA(15)	PGA(20)-Ub	0.63 ± 0.10

Fig. 4. Connectase (Cnt) catalyzes specific and efficient ligations without side reactions. (*A*) The time course of *M. mazei* Connectase and SrtA5*-mediated ligations of two ubiquitin molecules at various molar enzyme:substrate (1:1, 1:10, and 1:400) concentrations. Connectase is used at very low concentrations and therefore not visible on the gel. (*B*) Ligation rates for Connectase reactions with various peptide and protein substrate pairs at 25 µM. The values were determined in $n = 3$ independent experiments, with ± signifying the SD. (*C*) Michaelis–Menten plot of a *M. mazei* Connectase ligation using Ub-(5)KDPGA(15) and PGA(15)-Ub substrates. Although the Connectase reaction can be expected to follow more complex mathematics, the experimental data are well described by the Michaelis–Menten model (gray line). The values were determined in $n = 3$ independent experiments. (*D*) Polyacrylamide gel showing a test for *M. mazei* Connectase hydrolase activity with Ub-(5)KDPGA(15) and PGA(10) substrates. At an enzyme:substrate concentration of 1:1,000 (1×), Connectase catalyzes ∼50% product formation [i.e., Ub-(5)KDPGA(10)] in 30 min without forming the putative hydolysis product Ub-(5)KD. Even at 1,000× increased enzyme concentrations and prolonged incubation times, no hydrolysis product can be detected. (*E*) Polyacrylamide gel showing the ligation of two recombinantly expressed substrates [Strep-Ub-(5)KDPGA(10)-His_6 and PGA(15)-Ub] in cell lysates and the single-step purification of the respective ligation product [Strep-Ub-(5)KDPGA(15)-Ub] using an Ni-NTA column in series with a Streptavidin column.

good choice where short ligation motifs are preferred and high substrate concentrations available (40). Where these requirements are not met, Connectase is advantageous, as it combines substantially higher substrate affinities, reaction rates, and ligation yields without catalyzing detectable side reactions.

To study whether Connectase catalyzes such side reactions under more extreme conditions, we extended incubation times to several hours and increased Connectase concentrations by 1,000× (Fig. 4*D* and *SI Appendix*, Fig. S14). Over the entire period, the amount of ligation product remained constant at ∼50%, while no extra products could be detected. This finding may surprise, considering that all currently known ligase enzymes also function as proteases by forming an ester bond with the substrate that is then either transferred to water (hydrolysis/proteolysis) or to the amino group of a ligation partner (12). Connectase, however, differs from these enzymes in using the NTN active site architecture that allows a third alternative: the transfer to the N-terminal amino group and the associated formation of a hydrolysis-resistant amide-bonded intermediate (Fig. 3*C*).

To test whether the above characteristics hold true in more complex solutions, we ligated two independently expressed protein substrates, Strep-Ub-(5)KDPGA(10)-His_6 and PGA(15)-Ub within their respective cell lysate and subsequently purified the ligation products in a single step, using a Ni-NTA column in series with a streptavidin column (Fig. 4*E*). Only one protein species, Strep-Ub-(5)KDPGA(15), could be eluted with desthiobiotin, indicating that no other proteins in the lysate reacted with the Strep-Ub-(5)KD-Connectase intermediate. In accordance with the nanomolar affinity interaction between Connectase and its conserved recognition sequence (Fig. 2*B*), this observation suggests that Connectase ligations are highly specific and applicable even in complex solutions. Moreover, this experiment highlights the feasibility of Connectase-mediated ligations for the large-scale generation and single-step purification of a given ligation product in short time and with minimal amounts of enzyme.

Discussion

In Connectase we have identified a highly divergent proteasome homolog that, unlike all other members of the proteasome family, acts as a monomer and exclusively shows ligase activity. This activity throws an unexpected light onto a still poorly understood aspect of the proteasome: its splicing activity toward certain substrates (8). Here, the eukaryotic proteasome not only cleaves the protein chain into small peptide fragments but also recombines some of those fragments [1 to 2% (41)], which can then bind to major histocompatibility complex molecules and serve as antigens for the immune system. In contrast to Connectase ligations, proteasomal splicing is thought to involve no amide-bonded enzyme–substrate intermediate (Fig. 3 *C*, *Right*) and instead to function just by a reaction between the amino group of one peptide and a second ester-linked enzyme–substrate intermediate (i.e., the top left and middle reaction steps and intermediates in Fig. 3*C*). It is puzzling why proteasomal hydrolysis is disfavored in these reactions, given the high excess of water compared with transpeptidation substrates (42). Moreover, it is unclear to which extent this reaction is sequence-specific, yet the existence of S1′/S2′/S3′ pockets that bind the C-terminal splicing substrate has been postulated (41, 43)—in analogy to the S1/S2/S3 pockets which determine substrate specificity N-terminal of the cleavage site (44). Indeed, the (15)KDPGA(10)-Connectase structure shows the specific interaction of the C-terminal PGA(10) portion with what could be considered NTN domain S1′–S13′ pockets (Fig. 2*C*). Furthermore, the study of the Connectase reaction mechanism shows how ester-linked intermediates can be stabilized and thus escape hydrolysis by being transiently transferred to the active-site amino group (Fig. 3*C*). Thus, it may be worth studying in how far these aspects—binding mode and intermediate stabilization—are also realized in the proteasomal splicing reaction.

Connectase displays high specificity for a single substrate, MtrA. While the biological significance of this observation is the subject of ongoing study in our groups, we note that, in vitro, only a small fraction of MtrA is modified at any time (Fig. 1*A*). Therefore, it appears likely that $MtrA^{N}$-Connectase only serves as a reaction intermediate and that the biological function of Connectase lies in the recombination of two proteins bearing the substrate recognition sequence. Within MtrA, this recognition sequence separates the N-terminal catalytic domain from the C-terminal transmembrane anchor (23). Some methanogens encode several MtrA variants with highly similar catalytic domains that are coupled through gene conversion ((45), *SI Appendix*, Fig. S15). These paralogs differ in having either the canonical MtrA transmembrane anchor, the MtrF transmembrane anchor, the MtrG transmembrane anchor, or a soluble C-terminal fragment. Most of them retain the Connectase recognition sequence, but a recombination would result in very similar educts and products, due to their highly similar catalytic domains (up to 100%; *SI Appendix*, Fig. S15). Moreover, some methanogens do not encode for such extra MtrA variants and, based on profile searches with the MtrA interaction sites, we found no other candidates with a relevant recognition sequence in their genomes. Consequently, the physiological purpose of the Connectase reaction may instead be the exchange of identical MtrA catalytic domains with different modifications [e.g., cobalamin incorporation (23)] or associated proteins [e.g., different MtrH paralogs (45)]. In this case, the gene conversion events discussed above would not persist despite Connectase-mediated exchanges but because of them—to ensure equal functionality and exchangeability of the corresponding MtrA catalytic domains.

While the biological purpose of Connectase is still puzzling to us, we have made substantial progress in developing it as a tool for biotechnological applications. Here, it complements existing methods, such as native chemical ligation (46), expressed protein ligation (47), enzyme-mediated ligations (13), protein transsplicing (48, 49), or split domain systems (50). All of these methods are widely used, each with its functional niche due to its respective advantages and limitations (12, 51–54). While Connectase ligations require a longer linker between the ligated substrates than some of these methods, it also offers a unique combination of simplicity and efficiency: robust, fast, and highly specific ligations without side reactions that require no lengthy optimization procedures, complicated setup, special equipment, or defined buffer systems. Substrates can be produced recombinantly but are also amenable to chemical synthesis. With these characteristics, Connectase is highly attractive for a variety of applications, ranging from protein immobilization (e.g., on the cell surface, vesicles, or nanobeads) and fusion to chemical compounds (e.g., fluorophores and pharmacophores) to segmental isotope labeling and the generation of hybrid proteins (e.g., bispecific antibodies), especially upon optimization toward these specific tasks by directed evolution and protein design.

Methods

Bioinformatics. Connectase as was identified through HHpred (18, 55) searches with archaeal proteasome beta subunits against the Pfam database. The sequence alignment (*SI Appendix*, Fig. S1) was generated with PromalS3D (56), based on the crystallographic structures of the *M. jannaschii* proteasome beta subunit [Protein Data Bank (PDB) ID code 3H4P (57)] and *M. jannaschii* Connectase (*Crystal Structure Determination*). Sequence similarities were calculated based on all aligned residues (i.e., not counting the gaps), with similar amino acids defined by the following groups: GAVLI, FYW, CM, ST, KRH, DENQ, and P. The mapping of Connectase proteins on a phylogenetic tree (*SI Appendix*, Fig. S2) was done based on previously described phylogenetic realationships and metabolic analyses (58, 59).

Cloning, Expression, and Purification. Connectase genes MM_2909 and MJ_0548 as well MtrA genes MM_1543 and MJ_0851 were amplified via PCR from genomic DNA (DSM3647 and DSM2661 [DSMZ]), cloned into pET30 vectors and expressed recombinantly in *E. coli* BL21 DE3 cells (Stratagene) with the help of the rare codon plasmid pRARE. An exception presents a *M. jannaschii* Connectase variant with a removable N-terminal His_6-tag, which was used for crystallography and cloned into a pET28b vector instead. Through choice of appropriate PCR primers, the following variants were produced: *M. mazei* Connectase (MM_2909) was cloned together with a C-terminal His_6-(GSHHHHHH), Strep- (GSWSHPQFEK), Myc- (GSEQKLISEEDL) or HA-tag (GSYPYDVPDYA) and as active-site mutant $Connectase^{T1A}$ with C-terminal His_6-tag; *M. jannaschii* Connectase (MJ_0548) was cloned with a C-terminal His_6-tag, as active-site mutant $Connectase^{S1A}$ with C-terminal His_6-tag, without beta-barrel domain ($Connectase^{\Delta\beta B}$, lacking residues 203 to 293; C-terminal His_6-tag), as active-site mutant without beta-barrel domain ($Connectase^{\Delta\beta B\ S1A}$, C-terminal His_6-tag), and without NTN-domain ($Connectase^{\Delta NTN}$, lacking residues 1 to 200; C-terminal His_6-tag). Truncated MtrA constructs, *M. mazei* $MtrA^{\Delta 219–240}$ (MM_1543) with N-terminal Strep-tag (MWSHPQFEKGS) as well as *M. jannaschii* $MtrA^{\Delta 225–245}$ and $MtrA^{\Delta 174–245}$ (MJ_0851) with N-terminal His_6-tag (MHHHHHHGS), were also generated. In a similar way, based on the gene for mature *C. subterraneum* ubiquitin (Csub_C1474, synthesized by Eurofins), the following N-terminally His_6-tagged primary ligation substrates or C-terminally His_6-tagged secondary ligation substrates were produced: Ubiquitin fused C-terminally to the *M. mazei* MtrA residues 148 to 154 [Ub-(5)KD], 148 to 167 [Ub-(5)KDPGA(10)] or 148 to 172 [Ub-(5)KDPGA(15)] or N-terminally to a start-methionine plus *M. mazei* MtrA residues 155 to 167 [PGA(10)-Ub], 155 to 172 [PGA(15)-Ub] or 155 to 177 [PGA(20)-Ub]. Furthermore, the following genes were synthesized by Biocat: Mature *C. subterraneum* ubiquitin with N-terminal His_6-tag and C-terminal *M. mazei* MtrA 148 to 172 sequence plus Strep-tag [His_6-Ub-(5)KDPGA(15)-Strep], with N-terminal *M. mazei* MtrA 155 to 172 sequence and without affinity tag [PGA(15)-Ub], with N-terminal His_6-tag and C-terminal *M. jannaschii* MtrA sequence 154 to 173 [Ub- (5)KDPGA (10)], with C-terminal His_6-tag and N-terminal *M. jannaschii* MtrA sequence 161 to 178 [PGA(15)-Ub], with a C-terminal GGSLPETGGG modification and His_6-tag (Ub-GGSLPETGGG) or with an N-terminal His_6-tag following a TEV (tobacco etch virus) protease site for N-terminal glycine exposure (GGG-Ub). Camelid α-ricin single-domain antibody (60) (sdAb), CyP, or GST sequences were fused to an N-terminal *M. mazei* PGA(10) sequence and a C-terminal His_6-tag. Finally, SrtA and SrtA5* were synthesized as described previously (34), with an N-(SrtA) or C-terminal (SrtA5*) His_6-tag and without membrane anchor (i.e., without residues 1 to 59). All protein sequences and primers used for their generation are listed in Dataset S1.

E. coli cells were grown at 25 °C in M9 minimal medium supplemented with 50 μg/mL Se-Met, Leu, Ile, Phe, Thr, Lys, and Val for Se-Met labeling or in lysogeny broth (LB) for all other purposes. Protein expression was induced at an optical density of 0.4 at 600 nm with 500 μM isopropyl-β-D-thiogalactoside. Cells were harvested after 16 h, lysed by French press, and cleared from cell debris by ultracentrifugation. His_6-tagged proteins were purified via HisTrap HP columns (20 mM Tris·HCl, pH 8.0, 250 mM NaCl, 20 to 250 mM imidazole; all columns obtained from GE Healthcare) and Strep-tagged constructs via HiTrap Streptavidin HP columns (20 mM Hepes-NaOH, pH 7.5, 250 mM NaCl, 0 to 2.5 mM desthiobiotin). The His_6-tags of the GGG-Ub substrate and the N-terminally tagged Connectase variant for crystallography were then removed using the TEV or thrombin proteases, respectively. Next, the thermostable *M. jannaschii* and *C. subterraneum* proteins were incubated for 10 min at 80 °C and denatured protein removed via centrifugation. Finally, all proteins were applied to a Superdex 75 size-exclusion column (20 mM Hepes-NaOH, pH 7.5, 100 mM NaCl, 50 mM KCl, and 0.5 mM TCEP). All chromatography steps were performed on an Äkta Purifier FPLC (GE Healthcare) using Unicorn v5.1.0 software. Purified proteins were supplemented with 15% glycerol, flash-frozen in liquid nitrogen, and stored at −80 °C. Connectase and SrtA secondary substrates [PGA(X)-Ub and GGG-Ub] were subsequently analyzed via liquid chromatography mass spectrometry (LC-MS) (see *LC-MS*) to ensure complete removal of the start methionine.

Pulldown and Mass Spectrometrical Analysis. To obtain *M. mazei* cell extract, *M. mazei* strain Go1 was grown in a 100-L fermenter under standard growth conditions with 10 mM ammonium as nitrogen source and MeOH as sole energy and carbon source (61), cooled down, and harvested at exponential phase (2.8×10^8 cells/mL; optical density at 600 nm ∼1.6). A total of 20 g *M. mazei* cells were resuspended in 50 mL buffer A and lysed and insoluble fractions were removed via ultracentrifugation.

In a second step, *M. mazei* Connectase fused to Strep-, Myc-, or HA-tags was recombinantly expressed in 50 mL LB medium, lysed, and cleared from cell debris as described above, except for the use of a different resuspension buffer (buffer A: 20 mM MOPS-NaOH, pH 7.1, 150 mM NaCl, 100 mM KCl, and 0.15% Nonidet P-40). Magnetic Streptavidin (87.5 μL), anti-Myc, or anti-HA magnetic beads (175 μL each; Thermo) were then incubated for 1 h at room temperature with the above extracts containing Strep-, Myc-, or HA-tagged Connectase, respectively. Afterward, the beads were washed five times with 1 mL of buffer A and incubated with 15 mL *M. mazei* extract for 1 h at room temperature. After two washes with buffer A, bound proteins were eluted with 40 μL buffer A supplemented with 10 μM desthiobiotin or 2 mg/mL HA- or Myc-peptides, respectively. For mass spectrometrical analysis, bound proteins were separated via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), following an in-gel tryptic digest (13 ng/μL trypsin and 20 mM ammonium bicarbonate) (62). LC-MS/MS analysis was done on a Proxeon Easy nano-LC (Thermo) coupled to an LTQ OrbitrapElite mass spectrometer (Thermo). The data were processed using MaxQuant v1.6.4 (63) and spectra searched against the UniProt *M. mazei* Go1 proteome (*SI Appendix*, Fig. S3*A*).

Coelution Experiment. To verify the Connectase-MtrA interaction (Fig. 1*A*), equimolar (20 μM each in 1 mL) mixtures of either *M. mazei* Connectase-His_6 or Connectase^{T1A}-His_6 and Strep-MtrA$^{\Delta 219–240}$ were applied on a HisTrap HP 1-mL column (GE Healthcare). After rigorous washing with binding buffer (20 mM Tris·HCl, pH 8.0, 250 mM NaCl, and 20 mM imidazole), bound proteins were eluted in a single step with increased imidazole concentrations (250 mM) and the obtained fractions analyzed via SDS-PAGE.

LC-MS. For analysis of the MtrA-Connectase conjugate (Fig. 1 *B* and *C*), 0.5 g/L His_6-tagged *M. mazei* Connectase was incubated with 0.5 g/L Strep-tagged *M. mazei* MtrA$^{\Delta 219–240}$ in size-exclusion buffer for 20 min on ice. Similarly, for analysis of the Connectase-mediated protein–protein ligation (*SI Appendix*, Fig. S6), 0.5 g/L His_6-tagged *M. mazei* Connectase was incubated with 0.5 g/L His_6-tagged Ub-(5)KDPGA(10) and 0.5 g/L His_6-tagged PGA(10)-Ub. Desalted samples were subjected to a Phenomenex Aeris Widepore 3.6-μm C4 200-Å (100 × 2.1 mm) column, eluted with a 30 to 80% H_2O/acetonitrile gradient over 15 min in the presence of 0.05% trifluoroacetic acid, and analyzed with a Bruker Daltonik microTOF. Data processing was performed with Bruker Compass DataAnalysis 4.2 and the *m/z* deconvoluted with the MaxEnt module to obtain the protein mass.

Mass Spectrometrical Analysis of Dimethyl-Labeled MtrA-Connectase. To determine free amino groups in *M. mazei* Strep-MtrA$^{\Delta 219–240}$, Connectase-His_6, and their reaction product MtrA^N-Connectase, the corresponding bands were excised from an SDS gel similar to the one shown in Fig. 1*A*. Following an in-gel tryptic digest (13 ng/μL trypsin and 20 mM ammonium bicarbonate) (62), extracted protein fragments were desalted with C18 StageTips (64) and dimethylated [0.16% CH_2O, 22 mM $NaBH_3CN$, and 100 mM TEAB (65)] with an incorporation rate of 89 to 92%. Due to the use of different isotopes, this procedure resulted in 28-Da ("light," MtrA^N-Connectase), 32-Da ("medium," MtrA), or 36-Da ("heavy," Connectase) mass shifts per dimethylated amino group, respectively. LC-MS/MS analysis was performed subsequently on a Proxeon Easy nano-LC (Thermo) coupled to an LTQ OrbitrapElite mass spectrometer (Thermo). The data were processed using MaxQuant v.1.6.4 (63) and spectra searched against a custom peptide database and the UniProt *M. mazei* Go1 proteome (Fig. 1*D* and *SI Appendix*, Fig. S3*B*).

Light Scattering. Static light-scattering experiments (Fig. 2*A* and *SI Appendix*, Fig. S5 *A–D*) were performed with 50 μL 200 μM *M. jannaschii* Connectase^{S1A}-His_6, Connectase$^{\Delta\beta B\ S1A}$-His_6, Connectase$^{\Delta NTN}$-His_6, His_6-MtrA$^{\Delta 225–245}$, His_6-MtrA$^{\Delta 174–245}$, or 1:1 molar mixtures of the respective Connectase-MtrA pairs, using a Superdex S200 10/300 GL gel size-exclusion column (20 mM Hepes-NaOH, pH 7.5, 50 mM NaCl, and 100 mM KCl) coupled to a miniDAWN Tristar Laser photometer (Wyatt) and a RI-2031 differential refractometer (JASCO). Data analysis was carried out with ASTRA v7.3.0.18 software (Wyatt).

MST. For MST experiments (Fig. 2*B* and *SI Appendix*, Fig. S5 *E–L*), peptides based on the *M. jannaschii* MtrA sequence were synthesized by GenScript, with the fluorophore fluorescein isothiocyanate (FITC) either N-terminally linked via aminohexanoic acid [(15)KDPGA(5), (15)KDPGA(0)] or an extra C-terminal lysine [(15)KDPGA(10), (10)KDPGA(10), (5)KDPGA(10), (0)KDPGA(10); see *SI Appendix*]. Purified *M. jannaschii* His_6-MtrA$^{\Delta 225–245}$ and His_6-MtrA$^{\Delta 174–245}$ proteins were fluorescently labeled using the NT-647-NHS labeling kit (L001; Nanotemper). Next, a serial 1:2 dilution of *M. jannaschii* Connectase^{S1A} ranging from nano- to micromolar concentrations was prepared and mixed with 10 nM labeled peptide or 50 nM labeled protein (20 mM Hepes-NaOH, pH 7.5, 150 mM NaCl, 50 mM KCl, 0.5 mM TCEP, 0.05% Nonidet P-40, and 0.1 g/L bovine serum albumin). MST measurements were performed with a Monolith NT.115 (Nanotemper), using various MST power and laser intensity settings to test the general validity of the obtained data. The results were obtained in three independent experiments and measured at a temperature of 25 °C. The shown binding curve was fitted to the data, using the NT Analysis 1.5.41 software (Nanotemper).

Crystal Structure Determination. For initial crystallization experiments, *M. jannaschii* Connectase was concentrated to 15 mg/mL in 150 mM NaCl, 0.5 mM TCEP, and 20 mM Hepes pH 7.5. As a remnant of the Thrombin cleavage site (*Cloning, Expression, and Purification*), the employed protein variant had a four-residue modification, GSHM, on the N terminus. Screening of conditions was performed at 21 °C in 96-well sitting-drop vapor-diffusion plates with 600-nL drops containing equal volumes of protein solution and commercial screening solutions, equilibrated against a 50-μL reservoir. After optimization of conditions, best diffracting crystals were obtained with a crystallization buffer containing 70% MPD (2-methyl-2,4-pentanediol) and 100 mM Tris, pH 8.5. Crystals were loop-mounted directly from the plates and flash-cooled in liquid nitrogen. Diffraction experiments were performed at 100 K and a wavelength of 1 Å at beamline X10SA in 2009, using a MarCCD 225-mm charge-coupled device (CCD) detector. Data were indexed, integrated, and scaled using XDS (66), yielding a dataset in space group C2 with a resolution cutoff at 2.3 Å (*SI Appendix*, Table S1). Unit cell dimensions and space group led us to expect two to seven protein molecules in the asymmetric unit (ASU); attempts to solve the structure via molecular replacement using different truncated structures of proteasome beta subunits remained unsuccessful.

We thus reproduced the crystals using a selenomethionine-labeled version of the same *M. jannaschii* Connectase protein for anomalous dispersion experiments, which required further adjustment and optimization of conditions to obtain crystals of sufficient quality; best diffracting crystals were obtained at 30 °C with a protein concentration of 6 mg/mL and a crystallization buffer containing 70% MPD and 100 mM Hepes, pH 7.5, which were again loop-mounted directly from the plates and flash-cooled in liquid nitrogen. Due to the low lattice symmetry (C2) and the limited possibilities to record highly redundant data on the CCD detector, initial single-wavelength anomalous diffraction experiments at the Se K-edge yielded only faint anomalous difference signal. This led us to perform four-wavelength anomalous diffraction experiments, which yielded data to about 2.8 Å (*SI Appendix*, Table S1), with detectable anomalous differences extending to about 3.8 Å. After indexing, integration, and scaling using XDS, we successfully employed SHELXD (67) for heavy atom location, followed by substructure refinement using SHARP (68). After density modification with Solomon (69), the ARP/WARP secondary structure recognition pipeline (70) could trace the backbone of the four protein molecules in the asymmetric unit to large extents, such that most of the

structure could be built subsequently by Buccaneer (71). At this point, we switched to refine against the higher-resolution native data and completed the structure by cyclic manual modeling with Coot (72) and refinement with REFMAC5 (73) with local NCS restraints (*SI Appendix*, Table S1). The Ramachandran statistics (most favored/additionally allowed/generously allowed) are 94.6%/4.7%/0.4% as assessed with PROCHECK (74).

To obtain a complex structure, an equimolar ratio of Connectase^{S1A} and FITC-Ahx-(15)KDPGA(10) peptide was concentrated to 10.5 mg/mL in 50 mM NaCl, 0.5 mM TCEP, and 20 mM Hepes, pH 7.5. Crystallization screens were performed as described above. Best diffracting crystals were identified with condition #21 of the NeXtal ProComplex suite (100 mM Na-cacodylate and 15% PEG 4000) as crystallization buffer; prior to loop-mounting and flash-cooling in liquid nitrogen, crystals were briefly transferred to a droplet of crystallization buffer supplemented with 20% glycerol for cryoprotection. Diffraction experiments were performed at 100 K and a wavelength of 1 Å at beamline X10SA in 2019, using a Pilatus 6M-F hybrid pixel photon-counting detector. Data were indexed, integrated, and scaled using XDS, yielding a dataset in space group $P2_12_12_1$ with a resolution cutoff at 3.05 Å (*SI Appendix*, Table S1). Molecular replacement with MOLREP (75), using the above *M. jannaschii* Connectase structure as a search model, located two monomers in the ASU. After first rounds of refinement with REFMAC5, electron density for one peptide per protein monomer became apparent, which was built manually. The structure was completed by manual modeling with Coot and refinement with REFMAC5 with strong NCS restraints (*SI Appendix*, Table S1). According to PROCHECK, the Ramachandran statistics (most favored/additionally allowed/generously allowed) are 90.1%/9.2%/0.0%.

Activity Assays. Unless indicated otherwise, all reactions with *M. mazei* proteins were carried out in *M. mazei* reaction buffer (50 mM acetate, 50 mM MES, 50 mM Hepes, 150 mM NaCl, 50 mM KCl, and 5 mM TCEP, pH 7.0) and all experiments with *M. jannaschii* proteins conducted in *M. jannaschii* reaction buffer (50 mM MES, 200 mM NaCl, 50 mM KCl, and 5 mM TCEP, pH 5.8). The reactions were stopped at the indicated time points (see figures) by addition of 2% SDS and analyzed on Coomassie G-250 stained 12% NuPAGE SDS-polyacrylamide gels (Thermo).

The time course of the Connectase reaction with just one primary substrate (Fig. 3*B*, lanes 1 to 4) was performed by incubating 0.9 μM *M. mazei* Connectase-His_6 with 3.7 μM His_6-Ub-(5)KDPGA(10) at 37 °C. For the recombination experiment (Fig. 3*B*, lanes 5 to 10), 3.7 μM PGA(10)-sdAb-His_6, PGA(10)-CyP-His_6, or PGA(10)-GST-His_6 was added in addition and the samples incubated for 10 min before SDS-PAGE analysis. Time courses (*SI Appendix*, Fig. S7) of the above recombinations were performed in the same manner, except for the use of just 0.1 μM enzyme and for stopping the reaction at the indicated time points.

To investigate the effect of proteasome inhibitors on Connectase (*SI Appendix*, Fig. S8), 20 nM His_6-tagged *M. mazei* Connectase was incubated for 2 h at 25 °C with either 1% dimethyl sulfoxide (DMSO) or with 20 μM MG-132, *clasto*-Lactacystin β-Lactone or Epoxomicin (2 mM stocks dissolved in DMSO; Enzo Life Sciences). The activity of these samples was then determined at 50 °C by addition of 0.375 vol 13.3 μM His_6-Ub-(5)KDPGA(15) and PGA(15)-Ub-His_6 substrates for a final concentration of 5 μM substrates and 12.5 nM enzyme. To determine ligation rates with different recognition sequences (*SI Appendix*, Fig. S9), peptides based on the *M. mazei* MtrA sequence were synthesized by GenScript (Dataset S1). The experiments were then conducted by incubating 25 μM of the indicated primary and secondary substrates with 62.5 nM *M. mazei* Connectase-His_6 at 50 °C, following SDS-PAGE band quantification of the reaction time course. For the ligation with Ub-(5)KDPGA(15) and PGA(10)/PGA(15) peptides, an additional Strep-tag sequence (indicated by an asterisk) was added C-terminally of the Ub-(5)KDPGA(15) substrate, so that the ligation results in a size shift. After adjusting loaded protein quantities and the Coomassie staining procedure to ensure a linear, concentration-dependent signal increase, three independent experiments were conducted for each substrate combination. Protein bands were quantified with ImageJ 1.52a, assuming that all ubiquitin molecules bind the Coomassie dye in a similar manner. The obtained data showed that the observed product formation in reactions with equimolar educts correlated with the quotient between the current and maximum product formation:

$$Observed\ \ Ligation\ \ \% = \int_{t\_start}^{t\_end} [Maximum\ \ rate] * \left(1 - \frac{\text{Product concentration at t}}{\text{Maximum product concentration}}\right) * dt,$$

where the observed ligation % is the molar quotient between products and educts. This relationship was found to describe the empirical data best (see fits in *SI Appendix*, Figs. S9–S11) and used to approximate the maximum rate parameters given in Fig. 4*B*.

Kinetic parameters (Fig. 4*C* and *SI Appendix*, Fig. S10) of a Connectase reaction were determined by incubating the indicated concentrations of His_6-Ub-(5)KDPGA(15) and PGA(15)-Ub-His_6 at 50 °C with 400× lower concentrations of *M. mazei* Connectase-His_6. The maximum rate parameters of each experiment were determined as described above and used in a Michaelis–Menten plot. The kinetic parameters were determined based on a data fit.

The role of the C-terminal beta-barrel domain (*SI Appendix*, Fig. S11) was studied by recording and analyzing time courses of *M. jannaschii* Connectase or Connectase$^{\Delta\beta B}$ ligations as described above, except for the use of *M. jannaschii* MtrA-derived His_6-Ub-(5)KDPGA(10) and PGA(15)-Ub-His_6 substrates and a reaction temperature of 85 °C.

The reaction rates of *M. mazei* Connectase at different temperatures (*SI Appendix*, Fig. S12) were studied by incubating 6.25 μM His_6-Ub-(5)KDPGA(15) and PGA(15)-Ub-His_6 at either 25 °C, 37 °C, or 50 °C with 400× lower concentrations of *M. mazei* Connectase-His_6.

For a comparison with SrtA and SrtA5* (Fig. 4*A* and *SI Appendix*, Fig. S13), the data gathered in the kinetic analysis (discussed above) was compared with analogous SrtA5*/SrtA ligations using low (3.1 μM) or high (100 μM) substrate concentrations. The SrtA5*/SrtA reactions were conducted in the same manner as the Connectase reaction, except for the use of Ub-GGSLPETGGG-His_6 and GGG-Ub substrates, a reaction temperature of 37 °C, a different buffer system [50 mM Tris·HCl, pH 7.5, 150 mM NaCl, and 10 mM CaCl2 (76)] and higher enzyme concentrations as indicated. A data fit for SrtA5*/SrtA reactions was not possible due to the competing catalytic activities (hydrolysis and ligation).

A potential Connectase hydrolase activity (Fig. 4*D*) was investigated by incubating a mix of 25 μM His_6-Ub-(5)KDPGA(15)-Strep and 25 μM PGA(10) with either 25 nM or 25 μM *M. mazei* Connectase-His_6 at 50 °C, following SDS-PAGE analysis. In a related experiment, the potential hydrolase activity was determined without addition of PGA(10) (*SI Appendix*, Fig. S14).

For the ligation within crude extracts (Fig. 4*E*), Strep-Ub-(5)KDPGA(10)-His_6 and PGA(15)-Ub without affinity tag were recombinantly expressed in *E. coli*. The lysate of a 2-L culture was prepared as described above, except for using a different resuspension buffer (50 mM acetic acid, 50 mM MES, 50 mM Hepes, 100 mM NaCl, 50 mM KCl, 10 mM imidazole, 5 mM $MgCl_2$, 50 μg/mL DNase [Applichem], and cOmplete Protease Inhibitor [Roche], pH 7.0). For the ligation, equal volumes of each cell lysate were mixed with 0.09 g/L *M. mazei* Connectase-His_6, which corresponds to a molar enzyme:substrate ratio of roughly 1:30. After incubation for 15 min at 37 °C, the pH was then adjusted to 8.0 and the mixture applied on a HisTrap FF Ni^{2+}-NTA column in series with a HiTrap streptavidin column (GE). Following rigorous washing (20 mM Tris and 250 mM NaCl, pH 8), the reaction product was eluted with 2.5 mM desthiobiotin and, in a second step, His_6-tagged educts with 2.5 mM desthiobiotin and 250 mM imidazole.

Data Availability. Protein structure coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6ZVZ (wild-type Connectase) and 6ZW0 (complex of the Connectase S1A mutant with peptide). Nucleotide and amino acid sequences are provided in the *SI Appendix*. All study data are included in the article and the *SI Appendix*, and on the Mendeley public repository (doi:10.17632/m3yw9sf8nn.1).

ACKNOWLEDGMENTS. We thank Anja Rau, Lorena Vöhringer, and Eva Hertle (Max Planck Institute [MPI] for Developmental Biology) for technical assistance, and Victoria Sanchez (Core facility, MPI for Biochemistry, Martinsried) and Irina Droste-Borel (Proteome Center, University of Tübingen) for mass spectrometric analyses. We thank Kerstin Bär and Reinhard Albrecht for assistance with crystallographic sample preparation, Reinhard Albrecht and Christopher Heim for crystallographic data collection, and the staff of beamline X10SA/Swiss Light Source for excellent technical support. This work was supported by institutional funds from the Max Planck Society.

1. G. A. Collins, A. L. Goldberg, The logic of the 26S proteasome. *Cell* **169**, 792–806 (2017).
2. M. A. Humbard *et al.*, Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii. *Nature* **463**, 54–60 (2010).
3. M. J. Pearce, J. Mintseris, J. Ferreyra, S. P. Gygi, K. H. Darwin, Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. *Science* **322**, 1104–1107 (2008).
4. F. Striebel *et al.*, Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nat. Struct. Mol. Biol.* **16**, 647–651 (2009).
5. A. Lupas, J. M. Flanagan, T. Tamura, W. Baumeister, Self-compartmentalizing proteases. *Trends Biochem. Sci.* **22**, 399–404 (1997).
6. E. M. Huber *et al.*, A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nat. Commun.* **7**, 10900 (2016).

7. E. Di Cera, Serine proteases. *IUBMB Life* **61**, 510–515 (2009).
8. N. Vigneron *et al*., An antigenic peptide produced by peptide splicing in the proteasome. *Science* **304**, 587–590 (2004).
9. K. Morihara, Using proteases in peptide-synthesis. *Trends Biotechnol.* **5**, 164–170 (1987).
10. S. Liebscher *et al*., N-terminal protein modification by substrate-activated reverse proteolysis. *Angew. Chem. Int. Ed. Engl.* **53**, 3024–3028 (2014).
11. A. C. Braisted, J. K. Judice, J. A. Wells, Synthesis of proteins by subtiligase. *Methods Enzymol.* **289**, 298–313 (1997).
12. M. Schmidt, A. Toplak, P. J. Quaedflieg, T. Nuijens, Enzyme-mediated ligation technologies for peptides and proteins. *Curr. Opin. Chem. Biol.* **38**, 1–7 (2017).
13. S. K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763 (1999).
14. A. W. Jacobitz, M. D. Kattke, J. Wereszczynski, R. T. Clubb, Sortase transpeptidases: Structural biology and catalytic mechanism. *Adv. Protein Chem. Struct. Biol.* **109**, 223–264 (2017).
15. A. C. D. Fuchs, M. D. Hartmann, On the origins of symmetry and modularity in the proteasome family: Symmetry transitions are pivotal in the evolution and functional diversification of self-compartmentalizing proteases. *BioEssays* **41**, e1800237 (2019).
16. A. C. D. Fuchs *et al*., The architecture of the Anbu complex reflects an evolutionary intermediate at the origin of the proteasome system. *Structure* **25**, 834–845.e5 (2017).
17. A. C. D. Fuchs, L. Maldoner, K. Hipp, M. D. Hartmann, J. Martin, Structural characterization of the bacterial proteasome homolog BPH reveals a tetradecameric double-ring complex with unique inner cavity properties. *J. Biol. Chem.* **293**, 920–930 (2018).
18. J. Söding, Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
19. L. Zimmermann *et al*., A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
20. K. C. Costa, J. A. Leigh, Metabolic versatility in methanogens. *Curr. Opin. Biotechnol.* **29**, 70–75 (2014).
21. E. Bapteste, C. Brochier, Y. Boucher, Higher-level classification of the archaea: Evolution of methanogenesis and methanogens. *Archaea* **1**, 353–363 (2005).
22. Z. Lyu, Y. Lu, Metabolic shift at the class level sheds light on adaptation of methanogens to oxidative environments. *ISME J.* **12**, 411–423 (2018).
23. T. Wagner, U. Ermler, S. Shima, MtrA of the sodium ion pumping methyltransferase binds cobalamin in a unique mode. *Sci. Rep.* **6**, 28226 (2016).
24. S. Y. Jhan, L. J. Huang, T. F. Wang, H. H. Chou, S. H. Chen, Dimethyl labeling coupled with mass spectrometry for topographical characterization of primary amines on monoclonal antibodies. *Anal. Chem.* **89**, 4255–4263 (2017).
25. Y. Masforrol *et al*., Introducing an Asp-Pro linker in the synthesis of random one-bead-one-compound hexapeptide libraries compatible with ESI-MS analysis. *ACS Comb. Sci.* **14**, 145–149 (2012).
26. A. C. D. Fuchs, L. Maldoner, M. Wojtynek, M. D. Hartmann, J. Martin, Rpn11-mediated ubiquitin processing in an ancestral archaeal ubiquitination system. *Nat. Commun.* **9**, 2696 (2018).
27. Q. Xiao, F. Zhang, B. A. Nacev, J. O. Liu, D. Pei, Protein N-terminal processing: Substrate specificity of Escherichia coli and human methionine aminopeptidases. *Biochemistry* **49**, 5588–5599 (2010).
28. T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, Site-specific protein modification on living cells catalyzed by Sortase. *ChemBioChem* **9**, 802–807 (2008).
29. M. W. Popp, S. K. Dougan, T. Y. Chuang, E. Spooner, H. L. Ploegh, Sortase-catalyzed transformations that improve the properties of cytokines. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3169–3174 (2011).
30. T. Matsumoto, K. Furuta, T. Tanaka, A. Kondo, Sortase A-mediated metabolic enzyme ligation in Escherichia coli. *ACS Synth. Biol.* **5**, 1284–1289 (2016).
31. S. A. McConnell *et al*., Protein labeling via a specific lysine-isopeptide bond using the pilin polymerizing sortase from corynebacterium diphtheriae. *J. Am. Chem. Soc.* **140**, 8420–8423 (2018).
32. H. Hirakawa, S. Ishikawa, T. Nagamune, Design of Ca2+-independent Staphylococcus aureus sortase A mutants. *Biotechnol. Bioeng.* **109**, 2955–2961 (2012).
33. B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof, D. R. Liu, Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13343–13348 (2014).
34. I. Chen, B. M. Dorr, D. R. Liu, A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11399–11404 (2011).
35. T. Heck, P. H. Pham, A. Yerlikaya, L. Thony-Meyer, M. Richter, Sortase A catalyzed reaction pathways: A comparative study with six SrtA variants. *Catal. Sci. Technol.* **4**, 2946–2956 (2014).
36. J. M. Antos *et al*., Site-specific protein labeling via sortase-mediated transpeptidation. *Curr. Protoc. Protein Sci.* **89**, 15 13 11–15 13 19 (2017).
37. B. A. Frankel, R. G. Kruger, D. E. Robinson, N. L. Kelleher, D. G. McCafferty, Staphylococcus aureus sortase transpeptidase SrtA: Insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* **44**, 11188–11200 (2005).
38. J. Li *et al*., Optimization of sortase A ligation for flexible engineering of complex protein systems. *J. Biol. Chem.* **295**, 2664–2675 (2020).
39. D. A. Levary, R. Parthasarathy, E. T. Boder, M. E. Ackerman, Protein-protein fusion catalyzed by sortase A. *PLoS One* **6**, e18342 (2011).
40. J. M. Antos, M. C. Truttmann, H. L. Ploegh, Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118 (2016).
41. M. Mishto *et al*., Driving forces of proteasome-catalyzed peptide splicing in yeast and humans. *Mol. Cell. Proteomics* **11**, 1008–1023 (2012).
42. J. Liepe, H. Ovaa, M. Mishto, Why do proteases mess up with antigen presentation by re-shuffling antigen sequences? *Curr. Opin. Immunol.* **52**, 81–86 (2018).
43. N. Vigneron, V. Ferrari, V. Stroobant, J. Abi Habib, B. J. Van den Eynde, Peptide splicing by the proteasome. *J. Biol. Chem.* **292**, 21170–21179 (2017).
44. E. M. Huber *et al*., Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. *Cell* **148**, 727–738 (2012).
45. S. S. Wang, Y. H. Chen, Q. H. Cao, H. Q. Lou, Long-lasting gene conversion shapes the convergent evolution of the critical methanogenesis genes. *G3* **5**, 2475–2486 (2015).
46. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
47. T. W. Muir, D. Sondhi, P. A. Cole, Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6705–6710 (1998).
48. H. Wu, Z. Hu, X. Q. Liu, 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 (1998).
49. J. Zettler, V. Schütz, H. D. Mootz, The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).
50. B. Zakeri *et al*., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E690–E697 (2012).
51. G. T. Debelouchina, T. W. Muir, A molecular engineering toolbox for the structural biologist. *Q. Rev. Biophys.* **50**, e7 (2017).
52. S. S. Kulkarni, J. Sayers, B. Premdjee, R. J. Payne, Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nat. Rev. Chem.*, **2** (2018).
53. C. Sarmiento, J. A. Camarero, Biotechnological applications of protein splicing. *Curr. Protein Pept. Sci.* **20**, 408–424 (2019).
54. A. R. Sutherland, M. K. Alam, C. R. Geyer, Post-translational assembly of protein parts into complex devices by using SpyTag/SpyCatcher protein ligase. *ChemBioChem* **20**, 319–328 (2019).
55. V. Alva, S. Z. Nam, J. Söding, A. N. Lupas, The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44**, W410–W415 (2016).
56. J. Pei, M. Tang, N. V. Grishin, PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, W30–W34 (2008).
57. F. Zhang *et al*., Structural insights into the regulatory particle of the proteasome from Methanocaldococcus jannaschii. *Mol. Cell* **34**, 473–484 (2009).
58. I. Vanwonterghem *et al*., Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
59. P. N. Evans *et al*., Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
60. P. Legler *et al*., Stability of isolated antibody-antigen complexes as a predictive tool for selecting toxin neutralizing antibodies. *Abstr. Pap. Am. Chem. S*, **254** (2017).
61. K. Veit *et al*., Global transcriptional analysis of Methanosarcina mazei strain Gö1 under different nitrogen availabilities. *Mol. Genet. Genomics* **276**, 41–55 (2006).
62. N. Borchert *et al*., Proteogenomics of Pristionchus pacificus reveals distinct proteome structure of nematode models. *Genome Res.* **20**, 837–846 (2010).
63. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
64. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
65. P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A. J. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4**, 484–494 (2009).
66. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
67. G. M. Sheldrick, A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
68. C. Vonrhein, E. Blanc, P. Roversi, G. Bricogne, Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
69. J. P. Abrahams, A. G. W. Leslie, Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 30–42 (1996).
70. A. Perrakis, R. Morris, V. S. Lamzin, Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
71. K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
72. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
73. G. N. Murshudov, A. A. Vagin, A. Lebedev, K. S. Wilson, E. J. Dodson, Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 247–255 (1999).
74. R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton, PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
75. A. Vagin, A. Teplyakov, An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1622–1624 (2000).
76. C. S. Theile *et al*., Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1800–1807 (2013).