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

Adrian C. D. Fuchs*a,1, Moritz Ammelburg*b,1, Jörg Martin*a, Ruth A. Schmitza, Marcus D. Hartmann*a, and Andrei N. Lupasa,2

*aDepartment of Protein Evolution, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; and bInstitute 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)

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.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

PNAS 2021 Vol. 118 No. 11 e2017871118

https://doi.org/10.1073/pnas.2017871118

PNAS 2021 Vol. 118 No. 11 e2017871118

Published March 9, 2021.

Imprinted text: 0

1Present address: Fish & Richardson P.C., 80807 München, Germany.
2To whom correspondence may be addressed. Email: andrei.lupas@tuebingen.mpg.de.

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

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 H2O (hydrolysis), the proteasome and trypsin can also utilize the NH2 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).

Sequence-specific protein ligation is 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 ligation 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.
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 bioconjugation, 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 carboxyl 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 m ethionine 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 NT 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 years ago (22), the mesophilic M. m azei 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- or HA-tagged M. mazei Connectase as a bait. Mass spectrometrical 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. S3A). 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 (MtrAΔ219–240) with His₆-tagged Connectase from a Ni-NTA (nitritotriacetic acid) column (Fig. 1A). Surprisingly, this experiment also revealed the Thr₁-dependent formation of specific reaction products. Mass spectrometrical analysis identified these products as a C-terminal MtrA fragment (MtrA_C; Fig. 1B) and a conjugate between the corresponding N-terminal MtrA fragment and Connectase (MtrA_C–Connectase; Fig. 1C). This conjugate was 18 Da—that is, the mass of H₂O—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_S (i.e., the C-terminal MtrA fragment starting with PGA) and MtrAN–Connectase (i.e., the N-terminal MtrA fragment with KD fused to Connectase Thr-1). Accordingly, the tryptic fusion peptide (D-TLVIAFGK) that results from a conjugation of MtrA with the Connectase N terminus (TLVIAFGK...) 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. 1D). 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. S3B).

Connectase Recombines Proteins via the MtrA (5)KDPGA(10) Recognition Sequence. We investigated the structural basis for this reaction with the Connectase⁵⁰ active-site mutant and MtrA from the thermophilic M. jannaschii, as M. m azei 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. 2A and SI Appendix, Fig. S5 A and B). Microscale thermophoresis (MST) experiments with synthetic peptides containing the KDPGA motif plus N-terminal and C-terminal residues from the MtrA sequence, denoted as (x)KDPGA(y) in the following, show that the (0)KDPGA(10) sequence is sufficient for
for this high-affinity interaction (Fig. 2B and SI Appendix, Fig. S5 E–L).

We crystallized both Connectase alone and a (15)KDPGA(10)-Connectase\(^{\text{S1A}}\) complex and could solve their structures at a resolution of 2.3 Å and 3.05 Å, respectively (Fig. 2C and SI Appendix, Fig. S5M and Table S1). Compared with proteasome subunits, Connectase lacks a two-helix element mediating inter-subunit 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. S5M). 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 interactions 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, Caldariarchaeum 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. S3B, lanes 1 to 4). This suggested a reversible reaction, resulting in an equilibrium between modified and unmodified Ub-(5)KDPGA(10) (Fig. 3A). In this scenario, the “forward” reaction yields Ub-(5)KDC-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/CyP/GST resulted in equimolar ratios of educts and Ub-(5)KDPGA(10)-sdAb/CyP/GST products (Fig. 3B, 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. 1D), and the requirement of an active-site hydroxyl group as the catalytic nucleophile (Fig. L4), 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. 3C). Within the resulting equilibrium, most substrates (e.g., MtrA in Fig. L4) exist in the unmodified state (left side in Fig. 3C) and only a small fraction is bound to Connectase via an amide bond (right side in Fig. 3C). The hydrolysis-sensitive ester intermediate is almost not populated (SI Appendix, Fig. S3B), 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\(_2\) = 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. jan-naschii MtrA\(^{\text{D225E245}}\) and Connectase\(^{\text{S1A}}\) 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 \( \pm \) 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\(^{\text{S1A}}\), 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 \( \pm \) signifying the SD. (C) Crystal structures of M. jannaschii Connectase\(^{\text{S1A}}\) 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 SLBX (23)] could potentially bind to Connectase.

Fuchs et al.
Archaeal Connectase is a specific and efficient protein ligase related to proteasome β subunits
https://doi.org/10.1073/pnas.2017871118
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. 4A). The reaction is not inhibited by classical proteasome inhibitors, most likely because they cannot mimic the highly specific recognition sequence (SI Appendix, Fig. S8). A systematic analysis of sequence determinants N- and C-terminally of the MtrA-derived (x)KDGPAG(y)/PGA(z) motif suggests that 5 residues N-terminally of KDGPAG and 10 residues C-terminally of KDGPAG/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. 4B and SI Appendix, Fig. S9). Although these residues flanking the KDGPAG 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)KDGPAG(15) and PGA(15)-Ub was determined with 0.92 ligation per second and enzyme (catalytic constant \(k_{cat}\)). The half-maximum reaction speed was observed at a substrate concentration of \(\sim 2.2 \mu M\) each (Fig. 4C 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., *Methanococcus marisnigri*, \(T_{Opt} = 20 \pm 25\) °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 ligation reaction. Third, both Sortase and 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 Thr-1-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). Analagous to Connectase, SrtA ligates two sequences bearing the LPETG 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, Sortase also catalyzes the irreversible hydrolysis of substrates and products featuring this motif at a lower rate \([k_{cat}\text{Ligation}/k_{cat}\text{Hydrolysis} \approx 3.3 (37)]\) 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-GGSLPETG and GGG-Ub substrates and recorded the time course of the reaction. In our assays, SrtA5* displayed \(\sim 100\times\) higher ligase activity compared with SrtA but also catalyzed more 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-GGSLPETG and GGG-Ub substrates and recorded the time course of the reaction. In our assays, SrtA5* displayed \(\sim 100\times\) higher ligase activity compared with SrtA but also catalyzed more side reactions, possibly due to its low affinity for the secondary substrate \((K_M\text{LPETG} = 170 \mu M; K_M\text{GGG} = 4,700 \mu M)\) (34; SI Appendix, Fig. S13). Unlike Connectase, both Sortase variants showed only spurious ligation activity at low substrate concentrations (3 μM each) but displayed far more favorable characteristics at high substrate concentrations (100 μM each; Fig. 4A and SI Appendix, Fig. S13). Yet, even under those conditions, Connectase shows \(>4,000\times\) higher ligase activity than nonoptimized SrtA and \(>40\times\) increased ligation 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
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. 4D 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 ligation enzymes also function as proteases by forming an ester bond with the substrate that is then 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. 3C).

To test whether the above characteristics hold true in more complex solutions, we ligated two independently expressed protein substrates, Strep-Ub-(5)KDPGA(10)-His6 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. 4E). 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. 2B), 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 ligation, 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 MtrAN-Connectase only serves A only a small fraction of MtrA is modified at any time (Fig. 1 A). Thus, it may be worth studying in how far these aspects—binding mode and intermediate stabilization—are also realized in the proteasomal splicing reaction.

Methods

Bioinformatics. Connectase was identified through HHpred (18, 55) searches with archaeal proteasome beta subunits against the Pfam database (SI Appendix). Furthermore, we searched for homologs in the PromaSS3D (56), based on the crystallographic structures of the M. jannaschii proteasome beta subunit (Protein Data Bank [PDB] ID code 3HAP (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 relationships and metabolic analyses (58, 59).

Cloning, Expression, and Purification. Connectase genes MM_2909 and MJ_0548 as well as MtrA genes MM_1543 and MJ_0851 were amplified via PCR from genomic DNA (DMS3647 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 pRecAR (46). An exception presents a M. jannaschii Connectase variant with a removable N-terminal His6-tag, which was used for crystallization 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 His6-tag (GSHHHHHH), Strep- (GSWSHPQFEK), Myc- (GSEQKLISEEDL) or HA-tag (GSYPYDVPDYA) and as active-site mutant ConnectaseS1A with C-terminal His6-tag, without beta-barrel domain (ConnectaseS1A), lacking residues 203 to 293, C-terminal His6-tag, as active-site mutant without beta-barrel domain (ConnectaseS1A), C-terminal His6-tag, and without NTN-domain (ConnectaseS1N), lacking residues 1 to 200; C-terminal His6-tag. Truncated MtrA constructs, M. mazei MtrAΔβ (MM_1543) with N-terminal Strep-tag (GSWSHPQFEK) as well as M. jannaschii MtrAΔβ (MJ_0548) was cloned with a C-terminal His6-tag (MJHHHHHHS), 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 His6-tagged primary ligation substrates or C-terminally His6-tagged secondary ligation substrates were produced: Ubiquitin fused C-terminally to the M. mazei MtrA residues 148 to 154 [Ub-(5)KDGPA(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 His6-tag and C-terminal M. mazei MtrA residues 148 to 172 plus Strep-tag (His6-UB-(5)KDPGA(15)-Strep), with N-terminal M. mazei MtrA residues 155 to 172 sequenc- and without affinity tag [PGA(15)-Ub], with N-terminal His6-tag and C-terminal M. jannaschii MtrA sequence 154 to 173 [Ub-(5)KDPGA(10)], with C-terminal His6-tag and N-terminal M. jannaschii MtrA sequence 161 to 178 [PGA(15)-Ub], with a C-terminal GGLSPTGGMG modification and His6-tag (UB-GGLSPTGGMG) or with an N-terminal His6-tag following a TEV (tobacco etch virus) protease site for N-terminal glycine exposure (GGLG-Ub). Cameld α-ricin single-domain antibody (60), CYP, or GST sequences were fused to an N-terminal M. mazei PGA(10) sequence and a C-terminal His6-tag. Finally, α-ricin PTGS-RAS* were synthesized with an N-terminal His6-tag (α-ricin SrtA) or C-terminal (SrtRAS*) His6-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 containing 0.16% NaCl, 0.05 M tris(hydroxymethyl)aminomethane (Tris), and 0.05 M HCl with 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. His₂₅₉-tagged proteins were purified via HiTrap HP columns (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 20 to 250 mM imidazole; all columns obtained from GE Healthcare) and Streptagged constructs via HiTrap Streptavidin HP columns (20 mM Hepes-NaOH, pH 7.5, 250 mM NaCl, 0 to 2.5 mM desthiobiotin). The His₆-tags of the GGG-UB subunit and the N-terminally tagged Connectase variant for crystallography were then removed using the TEV or thrombin proteases, respectively. Next, the thermostable M. jannaschii 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 by 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 G01 was grown in a 100-L fermenter under standard growth conditions with 10 mM ammonium as nitrogen source and MoeH as sole energy and carbon source (61), cooled down, and harvested at exponential phase (2.8 × 10⁸ cells/mL; optical density at 600 nm −1.6). A total of 20 g of M. mazei cells were resuspended in 50 mL buffer A and lysed and insoluble fractions were removed by 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. After 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 and 2 mg/mL HA- or Myc-peptides, respectively. 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 v.1.6.4 (63) and spectra searched against a custom peptide database and the UniProt M. mazei G01 proteome (1.0 and SI Appendix, Fig. S3A).

Coelution Experiment. To verify the Connectase-MtrA interaction (Fig. 1A), equimolar (20 μM each in 1 mL) mixtures of either M. mazei Connectase-His₆ or Connectase with Strep-tagged M. mazei MtrA were applied to a HiTrap 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. 1B and C), 0.5 g/L His₆-tagged M. mazei Connectase was incubated with 0.5 g/L Strep-tagged M. mazei MtrA in 250 mM size-exclusion buffer for 20 min on ice. Similarly, for the analysis of the Connectase-mediated-protein–protein ligation (SI Appendix, Fig. S6), 0.5 g/L His₆-tagged M. mazei Connectase was incubated with 0.5 g/L His₆-tagged Ub (50 μM KDPA(10) and 0.5 g/L His₆-tagged PAGA(10)-Ub. Desalted samples were analyzed by liquid chromatography mass spectrometry (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 G01 proteome (1.0 and SI Appendix, Fig. S3A).

Light Scattering. Static light-scattering experiments (Fig. 2A and SI Appendix, Fig. S5 A–D) were performed with 50 μL 200 μM M. jannaschii Connectase-His₆, Connectase 225–245, His₆-Connectase 225–245, His₆-MtrA 225–245, His₆-MtrA 174–245, and 1:1 molar mixtures of the respective Connectase-MtrA pairs, using a Superdex 2600 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 v.7.3.0.18 software (Wyatt).

MST. For MST experiments (Fig. 2B and SI Appendix, Fig. S5 E–I), peptides based on the M. jannaschii MtrA sequence were synthesized by GenScript, with the fluorescein isothiocyanate (FITC) either C-terminal to SeMet or no SeMet amino acid ([15K]KDPA(5), (15K)DPGA(5)] or an extra C-terminal lysine ([15K]KDPA(10), (10K)DPGA(10), (5)KDPA(10), (0)KDPA(10); see SI Appendix, Figs. E–H, SI Appendix, Fig. S5 E–I) were fluorescently labeled using the NT-647-NHS labeling kit (L001; Nanotemper). Next, a serial 1.2 dilution of M. jannaschii Connectase (31) ranging from nano- to micromolar concentrations was mixed with 10 μM 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, which yielded data to about 2.8 Å (Fig. S5 B–D, Light Scattering), led us to expect two to seven protein molecules in the asymmetric unit (ASU); attempts to solve the structure via molecular replacement using different truncation 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 (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 P2₁ (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.
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 pH wasp modeling with Coot (72) and refinement with REFMAC (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 ConnectaseS1A 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 Heps, pH 7.5. Crystallization screens were performed as described above. Best diffusing crystals were identified with condition #21 of the NeXtal ProComex 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 crystalization 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 hybrid pixel photon-counting detector. Data were indexed, integrated, and scaled using XDS, yielding a dataset in space group P2₁2₁2₁ with a resolution cutoff at 3.05 Å (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 NaAc, 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. 38, lanes 1 to 4) was performed by incubating 0.9 μM M. mazei Connectase-His₆ with 3.7 μM His₆-Ub-(5)-KDPGA(10) at 37 °C. For the recombination experiment (Fig. 38, lanes 5 to 10), 3.7 μM PGA(10)-sdAb-His₆, PGA(10)-Cpy-His₆, or PGA(10)-GST-His₆ 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 His₆-tagged M. mazei Connectase was incubated for 2 h at 25 °C with either 1% dimethyl sulfoxide (DMSO) or with 20 μM MG-132, β-Lactacystin, MG-132, or Z-Leu-Leu-Leu-AMC (Bachem). The activity of these samples was then determined by incubating 6.25 μM His₆-Ub-(5)-KDPGA(15) and PGA(15)-Ubi-His₆ at either 25 °C, 37 °C, or 50 °C with 400× lower concentrations of M. mazei Connectase-His₆.

For a comparison with SrtA and SrtA* (Fig. 4A and SI Appendix, Fig. S5), the data gathered in the kinetic analysis (discussed above) was compared with analogous SrtA/SrtA* ligations using low (3.1 μM) or high (100 μM) substrate concentrations. The SrtASrta reactions were conducted in the same manner as the Connectase reaction, except for the use of Ub-GSGSLPETG-GG-His₆ 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 Srta*Srta reactions was not possible due to the competing catalytic activities (hydrolysis and ligation).

The reaction rates of M. mazei Connectase at different temperatures (SI Appendix, Fig. S12) were studied by incubating 6.25 μM His₆-Ub-(5)-KDPGA(15) and PGA(15)-Ubi-His₆ at either 25 °C, 37 °C, or 50 °C with 400× lower concentrations of M. mazei Connectase-His₆.

To investigate the effect of proteasome inhibitors on Connectase (SI Appendix, Fig. S8), 20 nM His₆-tagged M. mazei Connectase was incubated for 2 h at 25 °C with either 1% dimethyl sulfoxide (DMSO) or with 20 μM MG-132, β-Lactacystin, MG-132, or Z-Leu-Leu-Leu-AMC (Bachem). The activity of these samples was then determined by incubating 6.25 μM His₆-Ub-(5)-KDPGA(15) and PGA(15)-Ubi-His₆ at either 25 °C, 37 °C, or 50 °C with 400× lower concentrations of M. mazei Connectase-His₆.

For a comparison with SrtA and SrtA* (Fig. 4A and SI Appendix, Fig. S5), the data gathered in the kinetic analysis (discussed above) was compared with analogous SrtA/SrtA* ligations using low (3.1 μM) or high (100 μM) substrate concentrations. The SrtASrta reactions were conducted in the same manner as the Connectase reaction, except for the use of Ub-GSGSLPETG-GG-His₆ 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 Srta*Srta reactions was not possible due to the competing catalytic activities (hydrolysis and ligation).

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 ConnectaseS1B ligations as described above, except for the use of M. jannaschii Mtra-derived His₆-Ub-(5)-KDPGA(10) and PGA(15)-Ubi-His₆ 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 His₆-Ub-(5)-KDPGA(15) and PGA(15)-Ubi-His₆ at either 25 °C, 37 °C, or 50 °C with 400× lower concentrations of M. mazei Connectase-His₆.

For a comparison with SrtA and SrtA* (Fig. 4A and SI Appendix, Fig. S5), the data gathered in the kinetic analysis (discussed above) was compared with analogous SrtA/SrtA* ligations using low (3.1 μM) or high (100 μM) substrate concentrations. The SrtASrta reactions were conducted in the same manner as the Connectase reaction, except for the use of Ub-GSGSLPETG-GG-His₆ 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 Srta*Srta reactions was not possible due to the competing catalytic activities (hydrolysis and ligation).

For a comparison with SrtA and SrtA* (Fig. 4A and SI Appendix, Fig. S5), the data gathered in the kinetic analysis (discussed above) was compared with analogous SrtA/SrtA* ligations using low (3.1 μM) or high (100 μM) substrate concentrations. The SrtASrta reactions were conducted in the same manner as the Connectase reaction, except for the use of Ub-GSGSLPETG-GG-His₆ 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 Srta*Srta reactions was not possible due to the competing catalytic activities (hydrolysis and ligation).

For a comparison with SrtA and SrtA* (Fig. 4A and SI Appendix, Fig. S5), the data gathered in the kinetic analysis (discussed above) was compared with analogous SrtA/SrtA* ligations using low (3.1 μM) or high (100 μM) substrate concentrations. The SrtASrta reactions were conducted in the same manner as the Connectase reaction, except for the use of Ub-GSGSLPETG-GG-His₆ 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 Srta*Srta reactions was not possible due to the competing catalytic activities (hydrolysis and ligation).

For a comparison with SrtA and SrtA* (Fig. 4A and SI Appendix, Fig. S5), the data gathered in the kinetic analysis (discussed above) was compared with analogous SrtA/SrtA* ligations using low (3.1 μM) or high (100 μM) substrate concentrations. The SrtASrta reactions were conducted in the same manner as the Connectase reaction, except for the use of Ub-GSGSLPETG-GG-His₆ 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 Srta*Srta reactions was not possible due to the competing catalytic activities (hydrolysis and ligation).
11. A. C. Braisted, J. K. Judice, J. A. Wells, Synthesis of proteins by subtiligase.

23. T. Wagner, U. Ermler, S. Shima, MtrA of the sodium ion pumping methyltransferase.

13. S. K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, Staphylococcus aureus sortase.

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.

20. K. C. Costa, J. A. Leigh, Metabolic versatility in methanogens.

15. A. C. D. Fuchs, M. D. Hartmann, On the origins of symmetry and modularity in the intermediate at the origin of the proteasome system.

25. Y. Masforrol.

27. Q. Xiao, F. Zhang, B. A. Nacev, J. O. Liu, D. Pei, Protein N-terminal processing: Sub-cellular subunits.

28. T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, Site-specific protein modification on living cells catalyzed by Sortase.

32. H. Hirakawa, S. Ishikawa, T. Nagamune, Design of Ca2+-independent Staphylococcus protein systems.

33. B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof, D. R. Liu, Reprogramming the specificity of monoclonal antibodies.

34. G. T. Debeloulis, T. W. Muir, A molecular engineering toolbox for the structural biologist. Q. Rev. Biophys. 50, e7 (2017).

35. 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).

36. C. S. Theile et al., The architecture of the Anbu complex reflects an evolutionary transformation of methanogenesis and methanogens. Proc. Natl. Acad. Sci. U.S.A. 109, 8670–8675 (2012).

37. J. Zöld, Protein homology detection by HMM-HMM comparison. Bioinformatics 21, 951–960 (2005).

38. L. Zimmermann et al., A completely reimplemented MPI bioinformatics toolkit with a modern user interface at its core. J. Mol. Biol. 430, 2237–2243 (2016).

39. K. C. Costa, J. A. Leigh, Metabolic versatility in methanogens. Curr. Opin. Biotechnol. 30, 920–930 (2018).

40. J. Pei, M. Tang, N. V. Grishin, PROMALS3D web server for accurate multiple protein structure alignment. Nucleic Acids Res. 44, W410–W415 (2016).

41. M. Mishto et al., Driving forces of proteasome-catalyzed peptide splicing in yeast and human. Mol. Cell. Proteomics. 15, 1098–1103 (2012).

42. J. Liepe, H. Ovaa, M. Misho, Why do proteases mess up with antigen presentation by re-arranging their antigenic affinities? 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., Isolation of native proteasome crystal structures reveals differences in substrate and inhibitor specificity. Cell 158, 727–738 (2014).

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, Expression protein ligase: 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 DnaE gene of Synechocystis sp. PCC6803. Proc. Natl. Acad. Sci. U.S.A. 95, 9226–9231 (1998).

49. J. Zettler, V. Schütz, H. D. Moots, 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, E900–E907 (2012).

51. G. T. Debeloulis, 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. S. Theile et al., The architecture of the Anbu complex reflects an evolutionary transformation of methanogenesis and methanogens. Proc. Natl. Acad. Sci. U.S.A. 109, 8670–8675 (2012).

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. Zöld, A. N. Lupsa, The MPI bioinformatics Toolkit as an integrative platform for rational design of ligases. Science 350, 434–438 (2015).

56. P. Legler et al., Stability of isolated antibody-antigen complexes as a predictive tool for selecting toxin neutralizing antibodies. Abstr. Pap. Am. Chem. Soc. 254 (2017).

57. K. Veit et al., Global transcriptional analysis of Methanosarcina mazei strain G01 under different nitrogen availabilities. Mol. Genet. Genomics 276, 473–484 (2000).

58. I. Vanwonterghem et al., Mechanistic and structural insights into the acquired aryl-amine N-acetyltransferase discovery in the archaeal phylum Verrucomicrobia. Nat. Microbiol. 1, 16170 (2016).

59. P. N. Evans et al., Methane metabolism in the archael phylum Bathyrhaetaea 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. Soc. 254 (2017).

61. C. 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).

62. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-fragmentation, enrichment, pre-fragmentation and storage of peptides for proteomics using StageTip. Nat. Protoc. 2, 1896–1906 (2007).

63. P. J. Boersema, R. Rajmakers, S. Lemeer, S. Mohammed, A. J. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat. Protoc. 4, 484–494 (2009).

64. W. Kabsch, XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132 (2010).

65. G. M. Sheldon, A short history of SHELX. Acta Crystallogr. A 64, 112–122 (2008).

66. C. von Oheimb, E. Blanc, P. Roversi, G. Bricogne, Automated structure solution with autoSHARP. Methods Mol. Biol. 364, 215–230 (2007).

67. J. P. Abraham, A. G. W. Leslie, Methods used in the structure determination of bovine mitochondrial F1-ATPase. Acta Crystallogr. D Biol. Crystallogr. 52, 30–42 (1996).

68. A. Perrakis, R. Morris, V. S. Lamszín, Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. 6, 458–463 (1999).

69. K. Cowtan, The Buccaneer software for automated model building. 1. Tracking protein chains. Acta Crystallogr. D Biol. Crystallogr. 62, 1002–1011 (2006).

70. P. Emms, K. Cowtan, Coat: Model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 65, W34 (2008).

71. 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).

72. R. Recluz, W. M. Maier, D. S. Moss, J. M. Thornton, PROCHECK: A program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291 (1993).

73. A. Vagen, A. Tepljakov, An approach to multi-copy search in molecular replacement. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132 (2010).

74. C. S. Theile et al., Site-specific N-terminal labeling of proteins using sartate-mediated reactions. Nat. Protoc. 8, 1800–1807 (2013).