

Protein Macrocyclization for Tertiary Structure Stabilization

Anissa Haim,^[a] Saskia Neubacher,^[a, b] and Tom N. Grossmann^{*,[a, c]}

Proteins possess unique molecular recognition capabilities and enzymatic activities, features that are usually tied to a particular tertiary structure. To make use of proteins for biotechnological and biomedical purposes, it is often required to enforce their tertiary structure in order to ensure sufficient stability under the conditions inherent to the application of interest. The introduc-

tion of intramolecular crosslinks has proven efficient in stabilizing native protein folds. Herein, we give an overview of methods that allow the macrocyclization of expressed proteins, discussing involved reaction mechanisms and structural implications.

Introduction

Proteins possess unique folding characteristics and are central components of biological systems. The three-dimensional structure of folded proteins is determined by constraints of backbone flexibility and by intramolecular interactions that involve both backbone and side chain atoms.^[1] Usually, a particular tertiary structure is required for protein functionality, such as the engagement in molecular recognition processes or enzymatic activity. Proteins have been exploited for various biotechnological and biomedical applications, but their applicability is often hampered due to limited stability under the required conditions.^[2] For example, thermal and chemical stress can result in protein unfolding and consequently loss of function. To increase the usability of proteins, it is therefore of central importance to stabilize their tertiary structure.^[3] This has been achieved by optimizing intradomain interactions e.g. via directed evolution, consensus based mutagenesis, computational design or the introduction of unnatural amino acids.^[4–8]

In nature, Protein tertiary structures are additionally stabilized by disulfide bridges between two closely aligned cysteines,^[9] and in rare cases via head-to-tail cyclization.^[10] The resulting proteins possess cyclic topologies and often show a reduced tendency to unfold under thermal and chemical stress. In particular, the incorporation of additional disulfide bonds has been pursued to engineer proteins with sometimes considerably increased stability.^[9] However, disulfide engineering remains challenging since the short length of the cystine crosslink limits the number of suitable modification sites.^[9,11] This has

triggered the search for alternative protein macrocyclization approaches which, to some extent, utilize methodologies originally developed for the synthesis of cyclic peptides.^[12,13] Compared to peptides, the cyclization of proteins, however, possesses additional challenges which are mainly linked to the larger molecular weight of proteins and the difficulties to introduce unnatural amino acids. Both aspects complicate the design of sufficiently selective and reactive crosslinking reagents. Further, compared to peptides, the crosslinking of folded proteins requires more structural insight. Herein, we give an overview of strategies that have been used to cyclize expressed proteins in order to stabilize their tertiary structure. We present the mechanisms of involved cyclization reactions and discuss structural implications.


Disulfide Rebridging

Strategies that take advantage of existing disulfide bridges and convert them into chemically modified crosslinks usually aim at the introduction of additional functionalities and/or the generation of chemically inert structures.^[14,15] Efforts have particularly focused on the rebridging of inter-chain disulfides within antibodies^[16] utilizing bifunctional crosslinkers.^[17–21] The rebridging of intra-chain disulfides has been less pursued. In general, the first step involves the reduction of the existing disulfide (**1**, Figure 1a) usually by treatment with tris(2-carboxyethyl) phosphine (TCEP). The free cysteines (**2**) can then react with a bis-electrophilic reagent (**3**) which results in intramolecular crosslink formation to provide the rebridged protein (**4**, Figure 1a).^[22–24] One approach involves the use of an α,β -unsaturated- β^2 -monosulfone (**3a**, Figure 1b).^[22] Crosslinking proceeds initially via thiol addition to the monosulfone, after which sulfinic acid is eliminated to form a double bond. The second thiol addition then provides the final crosslink (**4a**). In the reported example, the monosulfone bis-electrophile bears a poly(ethylene glycol) (PEG) chain, which can promote protein biostability. This approach was applied for the cyclization of e.g. the protein ligand interferon- α 2b and the enzyme L-asparaginase, providing modified versions with retained tertiary structure and biological activity. In an alternative approach, 3,6-dichloro-1,2,4,5-tetrazine (**3b**, Figure 1b) was used as bis-

[a] A. Haim, Dr. S. Neubacher, Prof. Dr. T. N. Grossmann
Department of Chemistry and Pharmaceutical Sciences
VU University Amsterdam, Amsterdam (The Netherlands)
E-mail: t.n.grossmann@vu.nl

[b] Dr. S. Neubacher
InCircular B.V., De Boelelaan 1108
1081 HZ Amsterdam (The Netherlands)

[c] Prof. Dr. T. N. Grossmann
Amsterdam Institute of Molecular and Life Sciences
VU University Amsterdam, Amsterdam (The Netherlands)

 © 2021 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

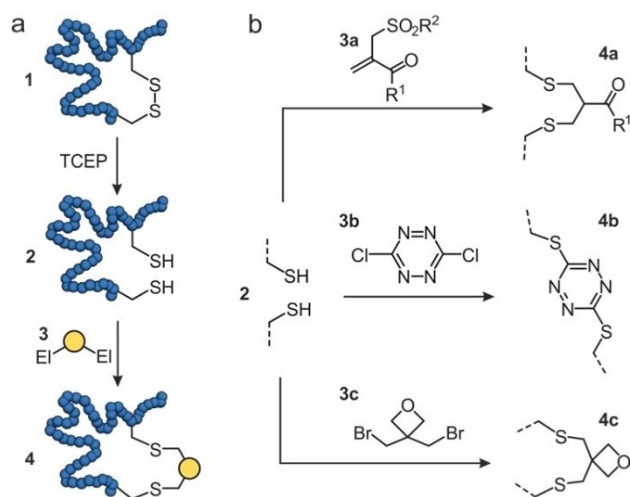


Figure 1. Schematic overview of disulfide rebridging: (a) A protein (blue) containing a disulfide (1) is reduced to thiols (2) by tris(2-carboxyethyl) phosphine (TCEP), followed by a reaction with a bis-electrophile (3) resulting in a rebridged protein (4). (b) Different bis-electrophiles have been reported: α,β -unsaturated- β -monosulfone (3a), 3,6-dichloro-1,2,4,5-tetrazine (3b), 3,3-bis(bromomethyl)oxetane (3c) to form new disulfide bridges 4a–4c, respectively.

electrophile allowing reactions at relatively low pH value.^[23] Tetrazine crosslinks (4b) have been introduced into a thioredoxin (Trx) protein and were shown to be photolytically removable. In another strategy, a 3,3-bis(bromomethyl)oxetane reagent (3c) was used to rebridge disulfide bonds (4c, Figure 1b) in detoxified diphtheria toxin CRM and the Trx protein.^[24] Rebridging of the thiols with 3c resulted in increased stability and

bioactivity. Notably for the diphtheria toxin, it was shown that an inaccessible disulfide within the protein core was not modified under applied conditions.

Head-to-Tail Cyclization

Head-to-tail (or backbone) cyclizations involve a lactam formation between the N- and C-terminus of a protein. In an early proof of principle, head-to-tail cyclization was achieved for a bovine pancreatic trypsin inhibitor by the activation of its C-terminus with a carbodiimide and subsequent intramolecular amide formation with the N-terminal amine.^[25] An alternative chemical cyclization approach involves the cyanylation of a C-terminal cysteine, thereby converting it into a leaving group and facilitating head-to-tail cyclization with the N-terminus.^[26] This reaction was applied to cyclize dihydrofolate reductase, which resulted in increased stability towards thermal and chemical denaturation. The generation of activated protein C-termini can also be accomplished using intein-based strategies. Inteins are protein domains that possess auto-cleavage activity and allow the fusion of their flanking N- and C-terminal extein sequences. This process is called protein cis-splicing and results in excision of the intein.^[27] The so-called expressed protein ligation (EPL) takes advantage of the ability of inteins to promote an N-to-S acyl transfer, thereby converting an amide bond (5, Figure 2a) into a thioester (6). Thereafter, reaction with a thiol promotes trans-thioesterification and cleavage of the intein sequence (8). The N-terminal cysteine then undergoes an intramolecular native chemical ligation with the C-terminal thioester to provide the head-to-tail cyclized protein (7, Fig-

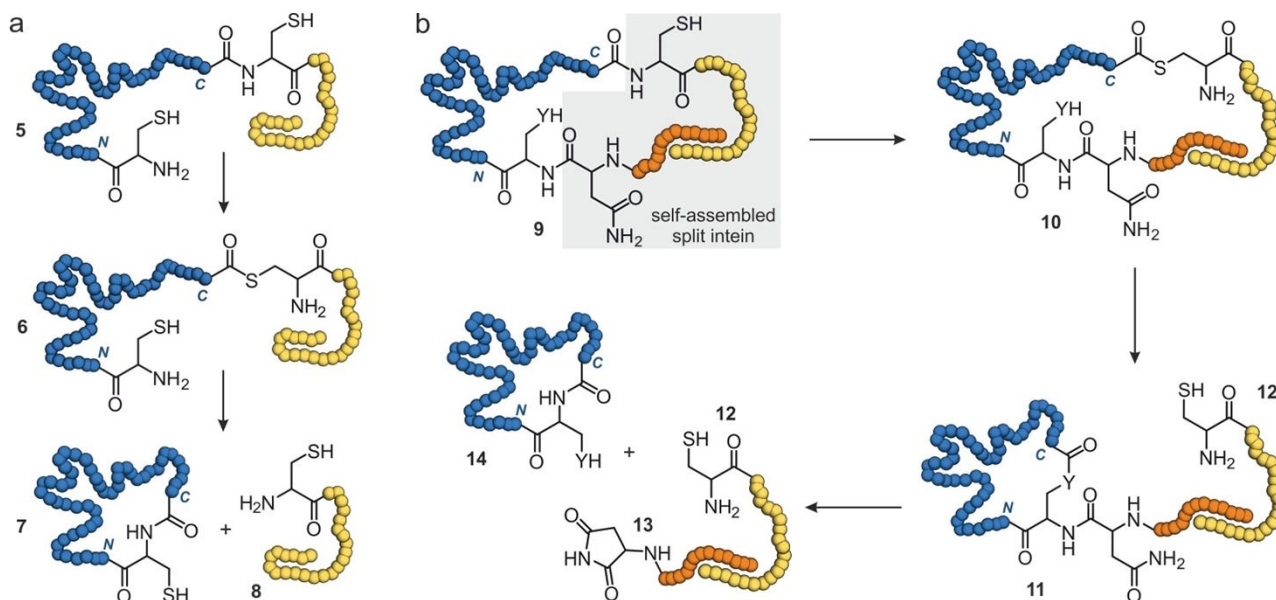


Figure 2. Schematic overview of intein-based head-to-tail cyclizations (blue: protein of interest): (a) Expressed protein ligation (EPL) involving N-to-S acyl transfer to convert an amide bond (5) to a thioester (6). This is followed by trans-thioesterification with a thiol to release the intein sequence (8, yellow) and subsequent intramolecular native chemical ligation to form the cyclic protein (7). (b) Split intein-based cyclization first involves assembly of the N- and C-terminal intein fragments (9, orange and yellow), after which an N-to-S acyl transfer takes place to form thioester 10. Next, trans-esterification with an intramolecular nucleophile (Y=S or O) occurs to form a lactone intermediate (11) to release the C-terminal intein fragment (12). Finally, Y-to-N acyl transfer cleaves the N-terminal intein (13) and forms cyclic protein 14.

ure 2a). An early example of EPL-mediated protein cyclization employed the Sce VMA intein for the cyclization of SH3 domains and of β -lactamase which, in most cases, resulted in proteins with increased stability.^[28–30] Using a different intein (Mxe GyrA), the effects of SH3 domain cyclization were analysed later in more detail revealing that cyclization-induced strain can also diminish the stabilizing effect.^[31]

The head-to-tail ligation in EPL-mediated cyclizations requires an N-terminal cysteine which then spontaneously attacks the formed thioester (6 to 7, Figure 2a). This can limit cyclization efficiency. An intein-based cyclization approach that supports the correct alignment of the N-terminus applies protein trans-splicing and is also called SICLOPPS (split intein circular ligation of peptides and proteins).^[32] This protein cyclization takes advantage of inteins that assemble from two fragments (split inteins), which are fused to the protein N- and C-terminus, respectively. If protein termini are located in proximity and if necessary appropriate linker sequences have been included, the split intein fragments assemble into the active splicing domain (9, Figure 2b). Similar to EPL, formation of a thioester (10) represents the initial step, which is then, however, followed by intramolecular trans-esterification with a particular cysteine or serine side chain ($Y=S$ or O). This results in the formation of a lactone intermediate (11) and release of the C-terminal intein sequence (12, Figure 2b). Subsequently, the N-terminal intein is cleaved under aspartimide (13) formation providing the cyclized protein (14).^[32] To date, several split inteins from different organisms have been used for protein cyclization. For example, DnaB and DnaE split inteins were retrieved from the cyanobacterium *Synechocystis* sp. *Strain PCC6803*.^[33,34] Early examples involve the use of the DnaE split intein for the cyclization of a dihydrofolate reductase and a maltose binding protein.^[32,35] Based on the DnaB intein, a mini-intein was engineered which has been utilized to cyclize various protein targets including the DnaB protein, a binary protein complex (FLINC4) and the enzyme Sortase A.^[36–38] In all three examples, the thermostability was improved. For cyclized Sortase A, increased enzymatic activity was observed under chemical stress (3 M urea). Another example, using the *PI-Pful* split intein from *Pyrococcus furiosus*, involves the cyclization of the green fluorescent protein (GFP) and of an endoglucanase both experiencing increased thermostability.^[39,40] Furthermore, a DnaE split intein was retrieved from *Nostoc Punctiforme* showing increased efficiency when compared to the DnaE intein from *Synechocystis* sp. *Strain PCC6803*.^[41] This split intein has been utilized to cyclize granulocyte colony-stimulating-factor thereby enhancing its thermostability.^[42]

Protein head-to-tail cyclization has also been performed using peptidases.^[43–45] For example, cysteine transpeptidases recognize their target sequence to initially cleave a peptide bond within the protein (15) which results in the formation of an activated thioester (17, Figure 3a). If aligned appropriately, the N-terminus of the target protein then attacks the thioester intramolecularly to provide the cyclic protein (18, Figure 3a). Transpeptidases used to generate cyclic proteins are Sortase A and Butelase 1. Sortase A from *Staphylococcus aureus* recognizes a five amino acid peptide sequence (L–P–aa–T–G, aa = any

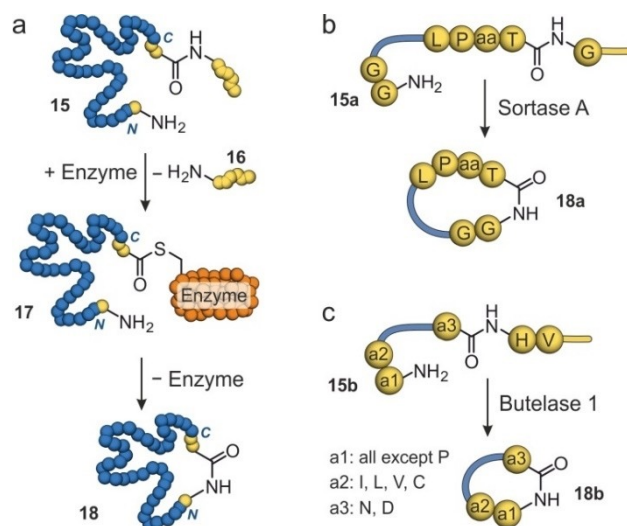


Figure 3. Schematic overview of enzymatic cyclization with Sortase A and Butelase 1: (a) The cysteine transpeptidase cleaves an amide bond in the recognition sequence of the protein (15, blue) forming a thioester intermediate (17) and releasing the C-terminal sequence (16, yellow). Subsequently, the reaction of the N-terminus with the thioester results in the cyclized protein (18). (b) Sortase A requires L–P–aa–T–G as the recognition motif and an oligoglycine at the N-terminus (15a) to generate a cyclic protein (18a). (c) Butelase 1 requires a3–H–V as the recognition motif and a1–a2 at the N-terminus (15b) to generate a cyclic protein (18b).

amino acid) and cleaves the amide between threonine (T) and glycine (G, 15a, Figure 3b). The N-terminus of the protein must contain an oligoglycine (15a) for cyclization to occur (18a, Figure 3b). Cyclization efforts with Sortase A were successful for different proteins, including Cre recombinase, GFP, human interferon α and granulocyte colony-stimulating-factor 3.^[46–48] More recently, the robustness of Sortase A has been improved, which can be expected to support the future application of Sortase A for head-to-tail cyclizations.^[49,50] Butelase 1 from *Clitoria ternatea* recognizes a short peptide sequence (a3–H–V; a3 = asparagine (N) or aspartic acid (D)) and cleaves the amide bond between position a3 and histidine (H, 15b, Figure 3c).^[51] To facilitate cyclizations, the first N-terminal amino acid (a1) can be any amino acid, except for proline (P), while the second amino acid (a2) must be isoleucine (I), leucine (L), valine (V) or cysteine (C, 15b, Figure 3c).^[52] Increased thermostability through Butelase 1 cyclization has been reported for an interleukin 1 receptor antagonist and a murine dihydrofolate reductase.^[51,53] It has been shown that cyclizations with Butelase 1 can occur with higher efficiency and faster kinetics, compared with Sortase A.^[51]

Isopeptide-Mediated Cyclization

The pilin structures from *Streptococcus pyogenes* (Spy0128) undergo spontaneous isopeptide bond formation between a lysine and an asparagine side chain.^[54] Both amino acids are embedded in a particular sequence context that facilitates the spatial alignment of a glutamic acid side chain which mediates

amide formation. Taking advantage of this ligation reaction, a protein labelling strategy has been developed, which was later also applied to perform protein cyclizations. In this setup, the Pilin-C (282 amino acids) and the Isopeptag (16 amino acids) sequences have been fused to the protein N- and C-terminus, respectively (Figure 4).^[55] In addition to the Pilin-C/Isopeptag pair, other lysine and asparagine-containing tags have been designed and used for protein cyclization, such as SnoopTag/SnoopCatcher (Figure 4).^[56,57] An alternative and frequently used pair of tags for protein cyclization involves the SpyTag (13 amino acids) and SpyCatcher (138 amino acids), originating from the collagen adhesin domain in *Streptococcus pyogenes*.^[58] SpyTag utilizes an aspartic acid which undergoes isopeptide bond formation with a lysine side chain in the SpyCatcher (Figure 4). Several proteins have been cyclized with the SpyTag fused to their N-terminus and SpyCatcher to the C-terminus of the protein resulting in proteins with increased thermostability.^[59–65] In some examples, SpyTag has been fused to the C-terminus and SpyCatcher to the N-terminus, also resulting in cyclic proteins with increased thermostability.^[66,67]

One study compares the performance of the three pairs Pilin-C/Isopeptag, SpyTag/Catcher and SnoopTag/Catcher for the cyclization of β -lactamase.^[59,60] In this study, cyclization with SpyTag/Catcher proceeded most efficiently and resulted in a β -lactamase version with reduced tendency for aggregation at high temperatures. Other studies comparing protein cyclization via SpyTag/Catcher and SnoopTag/Catcher also observed increased tolerance towards thermal stress for the SpyTag/Catcher pair.^[64,68] More recently, the RrgA domain of *Streptococcus pneumoniae* has been split into three domains providing SnoopTagJr (12 amino acids), DogTag (23 amino acids) and SnoopLigase (104 amino acids). SnoopTagJr and DogTag were fused to the N- and C-terminus of a protein, respectively (Figure 4). Only upon addition of SnoopLigase rapid protein cyclization was observed.^[57] Notably, for the two tested enzymes, β -lactamase and phytase C, the degree of stabilization with SnoopTagJr/DogTag exceeded what was earlier observed for the SpyTag/Catcher pair.^[59,60]

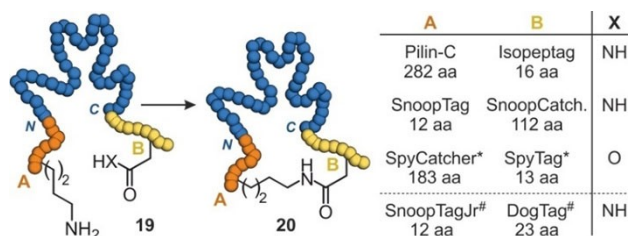


Figure 4. Isopeptide bond formation with different pairs of tags: (a) Protein (19, blue) with N- and C-terminal tags containing a lysine (A, orange) and an asparagine or aspartic acid (B, yellow), respectively, is cyclized (20). (b) Table with pairs of tags including the corresponding numbers of involved amino acids. * SpyCatcher and SpyTag have been fused to both N- and C-terminus. # Isopeptide formation requires the presence of SnoopLigase.

Unnatural Amino Acids

The introduction of unnatural amino acids (uaa) into proteins allows the installation of non-proteinogenic functionalities.^[69] One of the strategies to incorporate unnatural amino acids involves the suppression of the amber stop codon and its recognition by an orthogonal tRNA. This requires the availability of corresponding tRNA synthases and can come at the cost of lower expression yields.^[70] While early examples focused on the introduction of bioorthogonal reactive groups, it has also been employed to incorporate functionalities that allow intramolecular reactions with natural amino acids (Y, Figure 5a). For example, artificial disulfide bridges have been installed by the introduction of thiol-containing derivatives of tyrosine differing in the lengths of the thiol-alkyl chain (21 a, Figure 5b). Disulfide formation with a cysteine in spatial proximity then resulted in crosslink formation (22 a).^[71] This approach overcomes the geometrical constraints of natural cysteine disulfides, which only span relatively short distances. The introduction of this artificial disulfide in β -lactamase increased its thermal stability. Notably, the degree of stabilization was dependent on the length of the alkyl linker providing increases in the protein melting temperature (T_m) of up to 9°C.

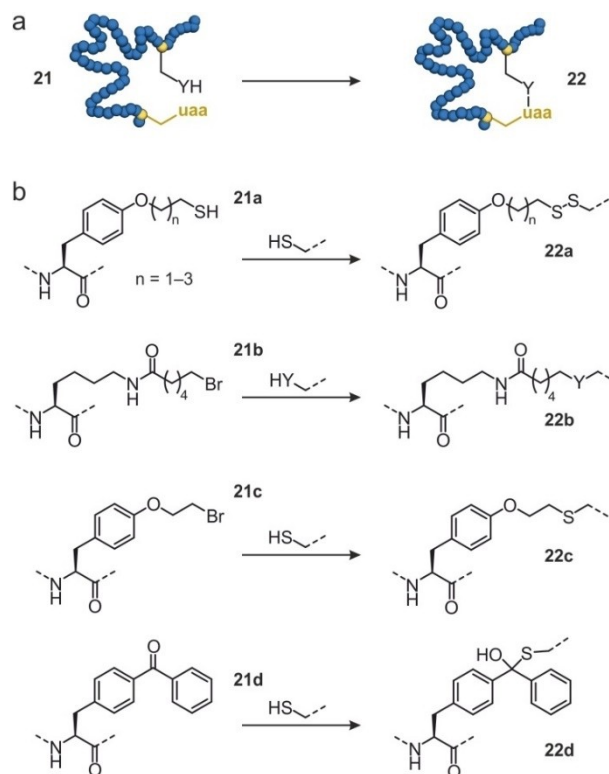


Figure 5. Protein cyclization utilizing unnatural amino acids: (a) Schematic overview of protein 21 with an unnatural amino acid (uaa) which undergoes an intramolecular reaction with a natural amino acid to form cyclic protein 22. (b) Overview of unnatural amino acids used for protein cyclization: O-(2-mercaptoalkyl)-L-tyrosine (21 a), (S)-2-amino-6-(6-bromohexanamido)-hexanoic acid (21 b), O-2-bromoethyl tyrosine (21 c) and (p-benzoylphenyl) alanine (21 d). They react with -SH (cysteine) or -YH (cysteine, histidine or lysine) to form the new bridges 22 a-d.

Alternatively, electrophilic unnatural amino acids have been introduced to undergo crosslinking reactions with nearby nucleophilic natural residues (**21 b-d**, Figure 5b).^[72–74] For instance, (S)-2-amino-6-(6-bromohexanamido)hexanoic acid (**21 b**, Figure 5b) was shown to react with cysteine, histidine and lysine (**22 b**).^[72] This approach was used for the cyclization of an affibody exhibiting high crosslinking efficiency, however at the cost of possible inhomogeneities in the cyclized product. Later, a computational approach was employed to identify potential crosslinking positions in a myoglobin-based cyclopropanation biocatalyst. After the incorporation of O-2-bromoethyl tyrosine (**21 c**, Figure 5b) and an appropriately aligned cysteine (**22 c**), crosslink formation was observed.^[73] Depending on the crosslink position within the protein, different stabilizing effects were observed (T_m increase of 4–10 °C). Notably, when combining two of these crosslinks, protein stability was further increased (T_m increase of 17 °C). In addition, enhanced stability towards chemical denaturation was observed while retaining catalytic activity and stereoselectivity.^[73] Intramolecular reaction with a cysteine side chain has also been achieved by incorporation of (p-benzoylphenyl)alanine (**21 d**, Figure 5b) resulting in the formation of a hemithioacetal crosslink (**22 d**). For a homoserine O-succinyltransferase, cyclization resulted in considerably increased thermal stability (T_m increase of 21 °C).^[74]

In Situ Cyclization

The site-specific crosslinking of proteins entirely composed of natural amino acids represents an appealing strategy. In an early example, 1,5-difluoro-2,4-dinitrobenzene was used to bridge a pair of solvent exposed lysine residues in ribonuclease A thereby increasing the protein's thermal stability.^[75,76] Due to the high frequency of lysine on protein surfaces and the resulting selectivity issues, later, the targeting of cysteines moved into the focus. In analogy to the bicyclization of peptides,^[77–79] C_3 -symmetric tris(bromomethyl)benzene was used to cyclize a *de novo* designed, artificial mini-protein exposing three cysteines.^[80] To facilitate the stabilization of natural tertiary structures, larger and water-soluble tris-electrophilic agents have been developed and used for the *in situ* cyclization of proteins (INCYPRO).^[50,81] Here, three surface-exposed cysteines are introduced (**23**, Figure 6a), preferably within different secondary structure elements.^[50,81] Initially, a hydrophilic tris-amine core (**Z1**) was decorated with chloroacetamide (**24 a**)^[82,83] and used to generate a bicyclic version of Sortase A. Notably, the INCYPRO reaction was selective for the three introduced cysteines tolerating the presence of an active-site cysteine.^[50] Bicyclic Sortase A exhibited increased thermal stability and performed labelling reactions under denaturing conditions. Later, a library of crosslinkers was composed using three different electrophiles (**24 a–24 c**, Figure 6b) which were combined with three different C_3 -symmetric core structures (**Z1–Z3**, Figure 6c).^[81] The resulting nine tris-electrophilic crosslinkers have then been used to cyclize the human KIX domain. All crosslinkers furnished bicyclic KIX domains with increased stability (T_m increase of 19–29 °C), with hydrophilic crosslink **Z1/**

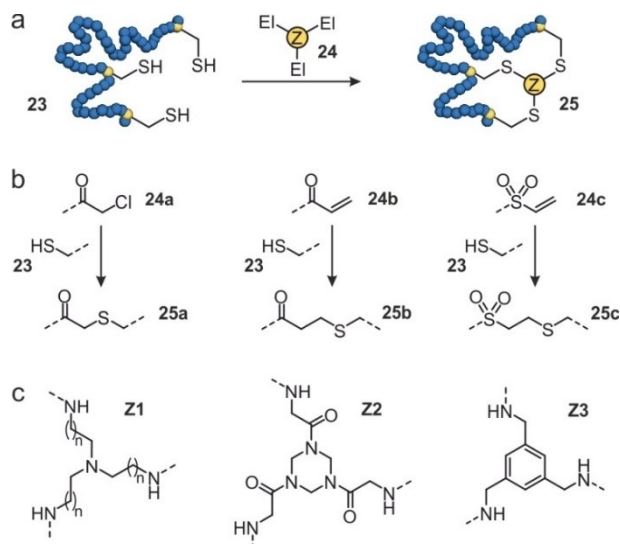


Figure 6. *In situ* cyclization of proteins (INCYPRO): (a) Schematic overview of protein (blue) with three cysteines (**23**) that can react with a tris-electrophile (**24**) to form a bicyclic protein (**25**). (b) Chemical structures of the electrophiles attached to the core of the tris-electrophile: chloroacetamide (**24 a**), acrylamide (**24 b**), and vinyl sulfonamide (**24 c**). They can react with cysteines to form **25 a–c**. (c) Chemical structures of the core of the tris-electrophile based on triethylamine (**Z1**), triazinane (**Z2**) and benzene (**Z3**).

25 a providing the most pronounced stabilizing effect. Interestingly, it was also found that the degree of protein stabilization correlated with the hydrophilicity of the crosslink.

Conclusions

To execute their biological function, proteins often require to adopt a folded state. For biotechnological and biomedical applications, it is usually necessary to stabilize the protein tertiary structure to prevent unfolding. Macrocyclization represents an efficient strategy towards the stabilization of protein tertiary structures. In fact, natural proteins already feature cyclic topologies, for example through the installation of disulfide bridges. Utilizing this feature, disulfide rebridging approaches have been pursued which usually do not aim for further tertiary structure stabilization, but for the selective introduction of additional functionalities or the generation of redox insensitive crosslinks. However, agents used for disulfide rebridging can also be expected to allow the crosslinking of newly installed cysteines. The remaining macrocyclization approaches described above aim for the installation of new intramolecular crosslinks and differ in particular in respect to i) the structural requirements for their introduction into the protein, ii) the level of modification caused within the protein, iii) the complexity of the methodology, and iv) the variability of the crosslink.

i) Structural requirements: All presented cyclization approaches require some degree of insight into the 3D structure of the protein. For example, head-to-tail cyclizations utilize protein termini that are located in proximity, and therefore in some cases linker sequences have to be introduced. Isopeptide

bond-mediated cyclizations are less sensitive to the orientation of the termini, as relatively large peptide tags are introduced. Crosslinking via unnatural amino acids or INCYPRO provides a large number of possible bridging sites, however, the protein structure or a precise homology model is usually required for the design process.

ii) Level of modification within the protein: With the exception of disulfide rebridging, all cyclization approaches require the variation of the protein sequence. The largest tags are introduced to facilitate isopeptide bond formation, while all other approaches usually only cause the variation or addition of a few amino acids. Intein-based as well as enzymatic cyclization approaches involve specific amino acid sequences at the ligation site and sometimes require the introduction of additional linker sequences to allow head-to-tail ligation.

iii) Complexity of the methodology: Cyclizations via split inteins, isopeptide bond-forming tags and unnatural amino acids mainly rely on biotechnological techniques required to prepare protein variants, while the actual crosslink formation proceeds spontaneously. The incorporation of unnatural amino acids for crosslinking involves additionally engineered cellular expression systems which tend to provide lower protein yields. The remaining approaches involve a separate crosslinking step.

iv) Variability of crosslink: The nature of the introduced crosslink can affect protein properties. For that reason, it may be advantageous to test various crosslink architectures also considering structures that go beyond the natural repertoire. Here, the use of unnatural amino acids and the INCYPRO approach are appealing. In particular, INCYPRO provides straight-forward access to diverse linker structures.^[81] In general, it can be expected that more cyclization strategies will be developed which allow the simultaneous introduction of an

additional functionality (e.g. PEG chains^[84] or photo-switchable moieties^[85,86]).

To compare structural prerequisites of different protein cyclization approaches, we selected β -lactamase and Sortase A as examples. Both are enzymes and therefore also illustrate effects of tertiary structure stabilization on enzymatic activity. β -Lactamases are single domain enzymes (Figure 7a) that support bacterial resistance towards β -lactam antibiotics.^[87] In an early example, a Sca VMA derived intein was used for EPL-mediated macrocyclization and resulted in the introduction of a 15 amino acid sequence between the original domain termini (Figure 7a, top).^[30] The cyclic enzyme exhibited a 5°C higher T_m -value, however, no activity tests at elevated temperatures were reported. Later a series of isopeptide-mediated cyclizations were described using various tag systems. Here, in particular, the refolding after thermal denaturation was studied revealing a lower tendency for aggregation for the cyclic versions.^[57,59,60] The smallest reported tag system (SnoopTagJr/DogTag, Figure 7a, middle) resulted in a β -lactamase version retaining solubility and activity following heat treatment up to 100°C.^[57] Screening a library of β -lactamases with deliberately positioned cysteines and randomly introduced thiol-containing unnatural amino acids provided an enzyme containing a disulfide between tyrosine derivative SbuY and a cysteine (Figure 7a, bottom). Its T_m -value was increased by 9°C when compared to the wildtype.^[71] Even though, the crosslink was located opposite to the enzyme's active site, the cyclic version showed lower activity than the wildtype under native conditions. At 40°C, however, the wildtype enzyme was inactive while the cyclic version retained its activity.

Sortase A is a single domain enzyme (Figure 7b) that catalyses transpeptidation reactions and has been applied for protein labelling. To increase its performance under chemical

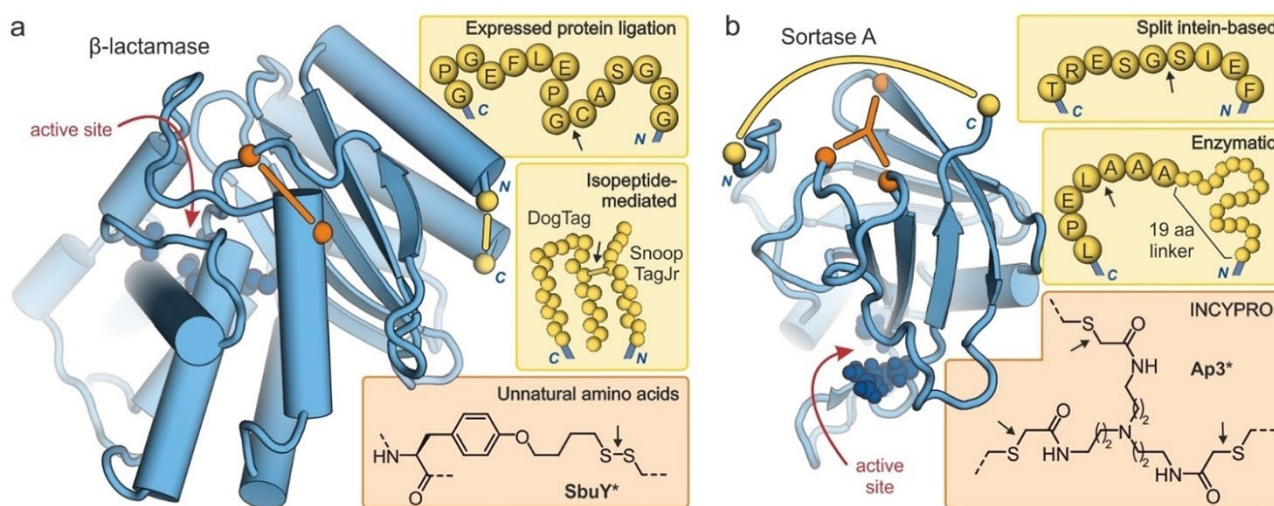


Figure 7. Structures of cyclized enzymes (blue spheres: active site residues, yellow: protein termini, orange: crosslinking sites). Arrows indicate the ligation site for macrocyclization: (a) Crystal structure of β -lactamase (TEM-1, PDB ID 1fqg)^[88] showing the introduced modifications resulting from macrocyclizations via EPL (top),^[30] isopeptide formation (middle, introduction of 23 and 12 amino acids respectively)^[57] and an unnatural amino acid (bottom, disulfide-bridge between cysteine and tyrosine derivative SbuY at positions 65 and 184, respectively).^[71] (b) NMR structure of Sortase A (PDB ID 1ija)^[89] showing the introduced modifications resulting from macrocyclizations via a split intein (top),^[38] enzymatic ligation (middle, 19-mer linker sequence: GSSHSHHHHSSGLVPRGSH),^[49] and INCYPRO (bottom, with Cys111, Cys149 and Cys177).^[50]

and thermal stress, different cyclization strategies have been pursued. Fusion with a DnaB split intein has provided a cyclic version with a 9-mer linker sequence (Figure 7b, top) exhibiting activity comparable to the wildtype enzyme under native conditions.^[38] Under denaturing conditions, cyclic Sortase A showed higher activity than the wildtype enzyme (e.g. in the presence of 2.5 M urea, 36% vs. 8% residual activity). Later, a more active Sortase version was head-to-tail cyclized using an enzymatic ligation.^[49] Here, a 26 amino acid sequence was introduced providing a cyclic enzyme (Figure 7b, middle) with slightly increased thermal stability ($\Delta T_m = 2^\circ\text{C}$) when compared with its linear precursor. However, under chemical stress, the cyclic version showed significantly increased resistance (e.g. in the presence of 1 M guanidine hydrochloride, 9% vs. <1% residual activity). The bicyclization of Sortase A using the INCYPRO approach (Ap3*, Figure 7b, bottom) resulted in increased thermal stability ($\Delta T_m = 11^\circ\text{C}$) and increased enzymatic activity above 55°C when compared to the linear version.^[50] Under denaturing conditions, bicyclic Sortase A showed considerably increased resistance (e.g. in the presence of 1 M guanidine hydrochloride, 36% vs. <1% residual activity).

Taken together, a variety of macrocyclization approaches is available, which may be selected based on the requirements of a particular application (e.g. straight-forward implementation, requirement for minimal modification or large crosslink diversity). Importantly, the meaningful stabilization of a protein often requires the introduction of multiple crosslinks.^[50,73] This renders the INCYPRO approach appealing which directly provides bicyclic proteins (when starting from linear precursors). Also, it should be highlighted that many of the presented cyclization approaches are orthogonal to each other and could in principle be combined to achieve high stabilization effects. The presented cyclic proteins have mainly found application as biocatalysts or ligands for cellular receptors. There is growing evidence that stabilized tertiary structures can also allow the inhibition of protein-protein interactions^[90–93] that are not addressable with classic peptidomimetic approaches.^[94,95] Therefore, it can be expected that cyclic proteins will find more fields of application in the future.

Acknowledgements

T.N.G. is grateful for support by the European Research Council (ERC starting grant number 678623).

Conflict of Interest

S.N. and T.N.G. are listed as inventors on a patent application related to the in situ cyclization of proteins, and they are co-founders of Incircular BV.

Keywords: disulfide mimetics · expressed protein ligation · INCYPRO · split inteins · unnatural amino acids

- [1] A. Sali, E. Shakhnovich, M. Karplus, *Nature* **1994**, *369*, 248–251.
- [2] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.
- [3] W. S. Horne, T. N. Grossmann, *Nat. Chem.* **2020**, *12*, 331–337.
- [4] M. T. Reetz, *Angew. Chem. Int. Ed.* **2013**, *52*, 2658–2666; *Angew. Chem.* **2013**, *125*, 2720–2666.
- [5] C. Jost, A. Plückthun, *Curr. Opin. Struct. Biol.* **2014**, *27*, 102–112.
- [6] T. J. Magliery, *Curr. Opin. Struct. Biol.* **2015**, *33*, 161–168.
- [7] A. M. Chapman, B. R. McNaughton, *Cell Chem. Biol.* **2016**, *23*, 543–553.
- [8] F. Agostini, J. S. Völler, B. Koks, C. G. Acevedo-Rocha, V. Kubyskhin, N. Budisa, *Angew. Chem. Int. Ed.* **2017**, *56*, 9680–9703; *Angew. Chem.* **2017**, *129*, 9810–9703.
- [9] A. A. Dombkowski, K. Z. Sultana, D. B. Craig, *FEBS Lett.* **2014**, *588*, 206–212.
- [10] L. Cascales, D. J. Craik, *Org. Biomol. Chem.* **2010**, *8*, 5035–5047.
- [11] V. S. Dani, C. Ramakrishnan, R. Varadarajan, *Protein Eng. Des. Sel.* **2003**, *16*, 187–193.
- [12] C. J. White, A. K. Yudin, *Nat. Chem.* **2011**, *3*, 509–524.
- [13] K. Shinbara, W. Liu, R. H. P. van Neer, T. Katoh, H. Suga, *Front. Chem.* **2020**, *8*, 447.
- [14] S. L. Kuan, T. Wang, T. Weil, *Chem. Eur. J.* **2016**, *22*, 17112–17129.
- [15] T. Lühmann, S. K. Mong, M. D. Simon, L. Meinel, B. L. Pentelute, *Org. Biomol. Chem.* **2016**, *14*, 3345–3349.
- [16] K. Yamada, Y. Ito, *ChemBioChem* **2019**, *20*, 2729–2737.
- [17] F. F. Schumacher, M. Nobles, C. P. Ryan, M. E. B. Smith, A. Tinker, S. Caddick, J. R. Baker, *Bioconjugate Chem.* **2011**, *22*, 132–136.
- [18] A. Maruani, M. E. B. Smith, E. Miranda, K. A. Chester, V. Chudasama, S. Caddick, *Nat. Commun.* **2015**, *6*, 6645.
- [19] J. P. M. Nunes, M. Morais, V. Vassileva, E. Robinson, V. S. Rajkumar, M. E. B. Smith, R. B. Pedley, S. Caddick, J. R. Baker, V. Chudasama, *Chem. Commun.* **2015**, *51*, 10624–10627.
- [20] S. Sun, P. Akkapeddi, M. C. Marques, N. Martínez-Sáez, V. M. Torres, C. Cordeiro, O. Boutureira, G. J. L. Bernardes, *Org. Biomol. Chem.* **2019**, *17*, 2005–2012.
- [21] A. J. Counsell, S. J. Walsh, N. S. Robertson, H. F. Sore, D. R. Spring, *Org. Biomol. Chem.* **2020**, *18*, 4739–4743.
- [22] S. Balan, J. W. Choi, A. Godwin, I. Teo, C. M. Laborde, S. Heidelberger, M. Zloh, S. Shaunak, S. Brocchini, *Bioconjugate Chem.* **2007**, *18*, 61–76.
- [23] S. P. Brown, A. B. Smith, *J. Am. Chem. Soc.* **2015**, *137*, 4034–4037.
- [24] N. Martínez-Sáez, S. Sun, D. Oldrini, P. Sormanni, O. Boutureira, F. Carboni, I. Compañón, M. J. Deery, M. Vendruscolo, F. Corzana, R. Adamo, G. J. L. Bernardes, *Angew. Chem. Int. Ed.* **2017**, *56*, 14963–14967; *Angew. Chem.* **2017**, *129*, 15159–14967.
- [25] D. J. Goldenberg, T. E. Creighton, *J. Mol. Biol.* **1983**, *165*, 407–413.
- [26] H. Takahashi, M. Arai, T. Takenawa, H. Sota, Q. H. Xie, M. Iwakura, *J. Biol. Chem.* **2007**, *282*, 9420–9429.
- [27] H. Paulus, *Chem. Soc. Rev.* **1998**, *27*, 375–386.
- [28] J. A. Camarero, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 5597–5598.
- [29] J. A. Camarero, D. Fushman, S. Sato, I. Gariat, D. Cowburn, D. P. Raleigh, T. W. Muir, *J. Mol. Biol.* **2001**, *308*, 1045–1062.
- [30] H. Iwai, A. Plückthun, *FEBS Lett.* **1999**, *459*, 166–172.
- [31] F. H. Schumann, R. Varadan, P. P. Tayakuniyil, J. H. Grossman, J. A. Camarero, D. Fushman, *Front. Chem.* **2015**, *3*, 1–15.
- [32] C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnson, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638–13643.
- [33] H. Wu, M. Q. Xu, X. Q. Liu, *Biochim. Biophys. Acta* **1998**, *1387*, 422–432.
- [34] H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226–9231.
- [35] T. C. Evans, D. Martin, R. Kolly, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X. Q. Liu, M. Q. Xu, *J. Biol. Chem.* **2000**, *275*, 9091–9094.
- [36] N. K. Williams, P. Prosser, E. Liepinsh, I. Line, A. Sharipo, D. R. Littler, P. M. G. Curmi, G. Otting, N. E. Dixon, *J. Biol. Chem.* **2002**, *277*, 7790–7798.
- [37] C. M. Jeffries, S. C. Graham, P. H. Stokes, C. A. Collyer, J. M. Guss, J. M. Matthews, *Protein Sci.* **2006**, *15*, 2612–2618.
- [38] D. Zhulenkova, K. Jaudzems, A. Zajacka, A. Leonchiks, *Biochem. Eng. J.* **2014**, *82*, 200–209.
- [39] H. Iwai, A. Lingel, A. Plückthun, *J. Biol. Chem.* **2001**, *276*, 16548–16554.
- [40] J. F. T. van Lieshout, O. N. P. Gutiérrez, W. Vroom, A. Planas, W. M. de Vos, J. van der Oost, S. Koutsopoulos, *Appl. Biochem. Biotechnol.* **2012**, *167*, 2039–2053.
- [41] H. Iwai, S. Züger, J. Jin, P. H. Tam, *FEBS Lett.* **2006**, *580*, 1853–1858.
- [42] T. Miyafusa, R. Shibuya, W. Nishima, R. Ohara, C. Yoshida, S. Honda, *ACS Chem. Biol.* **2017**, *12*, 2690–2696.
- [43] A. Purkayastha, T. J. Kang, *Biotechnol. Bioprocess Eng.* **2019**, *24*, 702–712.

- [44] T. Nuijens, A. Toplak, M. Schmidt, A. Ricci, W. Cabri, *Front. Chem.* **2019**, *7*, 829.
- [45] K. S. Harris, R. F. Guarino, R. S. Dissanayake, P. Quimbar, O. C. McCorkelle, S. Poon, Q. Kaas, T. Durek, E. K. Gilding, M. A. Jackson, D. J. Craik, N. L. van der Weerden, R. F. Anders, M. A. Anderson, *Sci. Rep.* **2019**, *9*, 10820.
- [46] J. M. Antos, M. W. L. Popp, R. Ernst, G. L. Chew, E. Spooner, H. L. Ploegh, *J. Biol. Chem.* **2009**, *284*, 16028–16036.
- [47] M. W. Popp, S. K. Dougan, T. Y. Chuang, E. Spooner, H. L. Ploegh, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3169–3174.
- [48] J. Hu, W. Zhao, Y. Gao, M. Sun, Y. Wei, H. Deng, W. Gao, *Biomaterials* **2015**, *47*, 13–19.
- [49] Z. Zou, D. M. Mate, M. Nöth, F. Jakob, U. Schwaneberg, *Chem. Eur. J.* **2020**, *26*, 13568–13572.
- [50] M. Pelay-Gimeno, T. Bange, S. Hennig, T. N. Grossmann, *Angew. Chem. Int. Ed.* **2018**, *57*, 11164–11170; *Angew. Chem.* **2018**, *130*, 11334–11170.
- [51] G. K. T. Nguyen, A. Kam, S. Loo, A. E. Jansson, L. X. Pan, J. P. Tam, *J. Am. Chem. Soc.* **2015**, *137*, 15398–15401.
- [52] G. K. T. Nguyen, Y. Qiu, Y. Cao, X. Hemu, C. F. Liu, J. P. Tam, *Nat. Protoc.* **2016**, *11*, 1977–1988.
- [53] X. Bi, J. Yin, X. Hemu, C. Rao, J. P. Tam, C. F. Liu, *Bioconjugate Chem.* **2018**, *29*, 2170–2175.
- [54] H. J. Kang, F. Coulibaly, F. Clow, T. Proft, E. N. Baker, *Science* **2007**, *318*, 1625–1628.
- [55] B. Zakeri, M. Howarth, *J. Am. Chem. Soc.* **2010**, *132*, 4526–4527.
- [56] G. Veggiani, T. Nakamura, M. D. Brenner, R. V. Gayet, J. Yan, C. V. Robinson, M. Howarth, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1202–1207.
- [57] C. M. Buldun, J. X. Jean, M. R. Bedford, M. Howarth, *J. Am. Chem. Soc.* **2018**, *140*, 3008–3018.
- [58] B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy, M. Howarth, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 690–697.
- [59] C. Schoene, J. O. Fierer, S. P. Bennett, M. Howarth, *Angew. Chem. Int. Ed.* **2014**, *53*, 6101–6104; *Angew. Chem.* **2014**, *126*, 6215–6104.
- [60] C. Schoene, S. P. Bennett, M. Howarth, *Sci. Rep.* **2016**, *6*, 21151.
- [61] J. Wang, Y. Wang, X. Wang, D. Zhang, S. Wu, G. Zhang, *Biotechnol. Biofuels* **2016**, *9*, 79.
- [62] Y. Wang, J. Tian, Y. Xiao, Y. Wang, H. Sun, Y. Chang, H. Luo, *Biotechnol. Lett.* **2019**, *41*, 987–994.
- [63] X. B. Sun, J. W. Cao, J. K. Wang, H. Z. Lin, D. Y. Gao, G. Y. Qian, Y. D. Park, Z. F. Chen, Q. Wang, *Nat. Biotechnol.* **2019**, *49*, 28–36.
- [64] D. Y. Gao, X. B. Sun, M. Q. Liu, Y. N. Liu, H. E. Zhang, X. L. Shi, Y. N. Li, J. K. Wang, S. J. Yin, Q. Wang, *J. Agric. Food Chem.* **2019**, *67*, 6837–6846.
- [65] Y. Wang, Y. Chang, R. Jia, H. Sun, J. W. Tian, H. Luo, H. Yu, Z. Shen, *Process Biochem.* **2020**, *95*, 260–268.
- [66] M. Si, Q. Xu, L. Jiang, H. Huang, *PLoS One* **2016**, *11*, 1–16.
- [67] C. Xu, Q. Xu, H. Huang, L. Jiang, *Biosci. Biotechnol. Biochem.* **2018**, *82*, 1473–1479.
- [68] Q. Xu, D. Ming, C. Shi, D. Lu, S. Tang, L. Jiang, H. Huang, *Process Biochem.* **2019**, *80*, 64–71.
- [69] H. Ren, *Biochem. Soc. Trans.* **2020**, *48*, 1807–1817.
- [70] T. S. Young, P. G. Schultz, *J. Biol. Chem.* **2010**, *285*, 11039–11044.
- [71] T. Liu, Y. Wang, X. Luo, J. Li, S. A. Reed, H. Xiao, T. S. Young, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5910–5915.
- [72] X. H. Chen, Z. Xiang, Y. S. Hu, V. K. Lacey, H. Cang, L. Wang, *ACS Chem. Biol.* **2014**, *9*, 1956–1961.
- [73] E. J. Moore, D. Zorine, W. A. Hansen, S. D. Khare, R. Fasan, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12472–12477.
- [74] J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, *J. Am. Chem. Soc.* **2018**, *140*, 15997–16000.
- [75] S. H. Lin, Y. Konishi, M. E. Denton, H. A. Scheraga, *Biochemistry* **1984**, *23*, 5504–5512.
- [76] P. C. Weber, S. Sheriff, D. H. Ohlendorf, B. C. Finzel, F. R. Salemme, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 8473–8477.
- [77] P. Timmerman, J. Beld, W. C. Puijk, R. H. Melen, *ChemBioChem* **2005**, *6*, 821–824.
- [78] C. Heinis, T. Rutherford, S. Freund, G. Winter, *Nat. Chem. Biol.* **2009**, *5*, 502–507.
- [79] S. Chen, D. Bertoldo, A. Angelini, F. Pojer, C. Heinis, *Angew. Chem. Int. Ed.* **2014**, *53*, 1602–1606; *Angew. Chem.* **2014**, *126*, 1628–1606.
- [80] B. Dang, H. Wu, V. K. Mulligan, M. Mravic, Y. Wu, T. Lemmin, A. Ford, D. A. Silva, D. Baker, W. F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10852–10857.
- [81] S. Neubacher, J. M. Saya, A. Amore, T. N. Grossmann, *J. Org. Chem.* **2020**, *85*, 1476–1483.
- [82] N. Brauckhoff, G. Hahne, J. T.-H. Yeh, T. N. Grossmann, *Angew. Chem. Int. Ed.* **2014**, *53*, 4337–4340; *Angew. Chem.* **2014**, *126*, 4425–4340.
- [83] C. Stiller, D. M. Krüger, N. Brauckhoff, M. Schmidt, P. Janning, H. Salamon, T. N. Grossmann, *ACS Chem. Biol.* **2017**, *12*, 504–509.
- [84] Q. Xiao, D. S. Ashton, Z. B. Jones, K. P. Thompson, J. L. Price, *RSC Chem. Biol.* **2020**, *1*, 273–280.
- [85] C. Hoppmann, V. K. Lacey, G. V. Louie, J. Wei, J. P. Noel, L. Wang, *Angew. Chem. Int. Ed.* **2014**, *53*, 3932–3936; *Angew. Chem.* **2014**, *126*, 4013–3936.
- [86] C. Hoppmann, I. Maslennikov, S. Choe, L. Wang, *J. Am. Chem. Soc.* **2015**, *137*, 11218–11221.
- [87] C. L. Tooke, P. Hinchliffe, E. C. Bragginton, C. K. Colenso, V. H. A. Hirvonen, Y. Takebayashi, J. Spencer, *J. Mol. Biol.* **2019**, *431*, 3472–3500.
- [88] N. C. J. Strynadka, H. Adachi, S. E. Jensen, K. Johns, A. Sielecki, C. Betzel, K. Sutoh, M. N. G. James, *Nature* **1992**, *359*, 700–705.
- [89] U. Ilangovan, H. Ton-That, J. Iwahara, O. Schneewind, R. T. Clubb, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6056–6061.
- [90] J. Sadek, M. G. Wuo, D. Rooklin, A. Hauenstein, S. H. Hong, A. Gautam, H. Wu, Y. Zhang, E. Cesarman, P. S. Arora, *Nat. Commun.* **2020**, *11*, 1786.
- [91] H. Adihou, R. Gopalakrishnan, T. Förster, S. M. Guéret, R. Gasper, S. Geschwindner, C. Carrillo García, H. Karatas, A. V. Pobbati, M. Vazquez-Chantada, P. Davey, C. M. Wassvik, J. K. S. Pang, B. S. Soh, W. Hong, E. Chiarparin, D. Schade, A. T. Plowright, E. Valeur, M. Lemurell, T. N. Grossmann, H. Waldmann, *Nat. Commun.* **2020**, *11*, 5425.
- [92] S. H. Hong, D. Y. Yoo, L. Conway, K. C. Richards-Corke, C. G. Parker, P. S. Arora, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2101027118.
- [93] M. Wendt, R. Bellavita, A. Gerber, N.-L. Efrém, T. van Ramshorst, N. M. Pearce, P. R. J. Davey, I. Everard, M. Vazquez-Chantada, E. Chiarparin, P. Grieco, S. Hennig, T. N. Grossmann, *Angew. Chem. Int. Ed.* **2021**, *60*, 13937–13944; *Angew. Chem.* **2021**, *133*, 14056–14063.
- [94] M. Pelay-Gimeno, A. Glas, O. Koch, T. N. Grossmann, *Angew. Chem. Int. Ed.* **2015**, *54*, 8896–8927; *Angew. Chem.* **2015**, *127*, 9022–8927.
- [95] A. D. Cunningham, N. Qvit, D. Mochly-Rosen, *Curr. Opin. Struct. Biol.* **2017**, *44*, 59–66.

Manuscript received: March 11, 2021
Revised manuscript received: May 26, 2021
Accepted manuscript online: May 31, 2021
Version of record online: June 21, 2021