Fan, Ruxia; Hakanpää, Johanna; Elfving, Karoliina; Taberman, Helena; Linder, Markus B.; Aranko, A. Sesilja

Biomolecular Click Reactions Using a Minimal pH-Activated Catcher/Tag Pair for Producing Native-Sized Spider-Silk Proteins

Published in:
Angewandte Chemie - International Edition

DOI:
10.1002/anie.202216371

Published: 06/03/2023

Document Version
Publisher's PDF, also known as Version of record

How to cite: *Angew. Chem. Int. Ed.* **2023,** *62,* e202216371
doi.org/10.1002/anie.202216371

VIP **Protein Engineering** **Very Important Paper**

Biomolecular Click Reactions Using a Minimal pH-Activated Catcher/Tag Pair for Producing Native-Sized Spider-Silk Proteins**

Ruxia Fan, Johanna Hakanpää, Karoliina Elfving, Helena Taberman, Markus B. Linder, and A. Sesilja Aranko**

Abstract: A type of protein/peptide pair known as Catcher/Tag pair spontaneously forms an intermolecular isopeptide bond which can be applied for biomolecular click reactions. Covalent protein conjugation using Catcher/Tag pairs has turned out to be a valuable tool in biotechnology and biomedicines, but it is essential to increase the current toolbox of orthogonal Catcher/Tag pairs to expand the range of applications further, for example, for controlled multiple-fragment ligation. We report here the engineering of novel Catcher/Tag pairs for protein ligation, aided by a crystal structure of a minimal CnaB domain from *Lactobacillus plantarum.* We show that a newly engineered pair, called Silk-Catcher/Tag enables efficient pH-inducible protein ligation in addition to being compatible with the widely used SpyCatcher/Tag pair. Finally, we demonstrate the use of the SilkCatcher/Tag pair in the production of native-sized highly repetitive spider-silk-like proteins with >90 % purity, which is not possible by traditional recombinant production methods.

Introduction

Advances in the methods of covalent ligation of proteins in vivo and in vitro have expanded the possibilities of recombinant production and created numerous applications, such as engineering non-linear architectures, as well as protein semisynthesis, cyclization, and purification.[1–5] Protein ligation using Catcher/Tag pairs has received wide attention since the first reports in 2010 due to the efficiency and robustness of the reaction.[6–8] The method is based on a biomolecular "click" reaction in which an autocatalytic isopeptide bond covalently links a Catcher protein and a Tag peptide upon protein fragment complementation. The Catcher/Tag pairs are derived from CnaB domains, Ig-like domains found in collagen-binding cell surface proteins in gram-positive bacteria, which carry intramolecular isopeptide bonds stabilizing the structure.[9,10] Some CnaB domains can be split into two fragments: a ≈10 kDa larger fragment, named Catcher, and a ≈1.5–2.5 kDa peptide, called Tag, which associate by protein fragment reconstitution. The association is followed by an autocatalytic isopeptide-bond formation, thus leading to covalent conjugation of the protein-peptide pair and proteins fused to them. Elegant engineering combined with directed evolution has further improved the SpyCatcher/Tag pair, turning it into a powerful tool for the covalent conjugation of proteins which has been adapted for multiple applications.[4,11–13]

A widely interesting application of protein conjugation is multiple fragment ligation[14–16] that allows the recombinant production of long and repetitive proteins that are difficult to produce in the common production hosts, such as *Escherichia coli.* Spider silk is an important and interesting example of this, because of its extreme properties at both the molecular and macroscopic levels, as well as the numerous applications.[17–19] The desirable mechanical properties of spider-silk fibers, combined with the unfeasibility of farming spiders, have prompted research aiming at producing recombinant spider silk. Spider silk proteins, called spidroins, consist of very long (thousands of amino acids, 250–320 kDa) and highly repetitive alanine- and glycine-rich regions.[20,21] The size and repetitive nature of the core parts are crucial for the fiber properties, but they create significant challenges for the recombinant production of the spidroins.[18,19]

Multiple fragment ligation is, however, limited by the lack of efficient and orthogonal (non-cross-reactive) protein ligation methods. The use of non-orthogonal methods would likely lead to predominantly intramolecular cyclization reactions.[22–24] In addition, the length of the products of intermolecular conjugation could not be controlled but instead multimers of varying lengths would be formed. The uncontrolled formation of side products makes the system

[*] R. Fan, K. Elfving, Prof. M. B. Linder, Dr. A. S. Aranko
Department of Bioproducts and Biosystems,
School of Chemical Engineering, Aalto University
P.O. Box 16100, 02150 Espoo (Finland)
E-mail: markus.linder@aalto.fi
sesilja.aranko@aalto.fi
Dr. J. Hakanpää, Dr. H. Taberman
Deutsches Elektronen Synchrotron (DESY), Photon Science
Notkestrasse 85, 22607 Hamburg (Germany)
Dr. J. Hakanpää
Hamburg Unit c/o DESY,
European Molecular Biology Laboratory (EMBL)
Notkestrasse 85, 22603 Hamburg (Germany)

[**] A previous version of this manuscript has been deposited on a preprint server (https://doi.org/10.1101/2022.10.21.513168).

unsuitable for any larger scale production. Thus, a toolbox of orthogonal Catcher/Tag pairs is needed. In addition to being orthogonal, these pairs should preferably work in the same conditions and be robust in terms of the target proteins. The successful engineering of SpyCatcher/Tag pair has inspired search for other Catcher/Tag pairs.[14,25–27] The compatibility of these pairs with SpyCatcher/Tag pair is, however, limited due to limited efficiency,[27] incompatible pH range preventing one-pot reactions,[14,25] and cross-reactivity.[26]

In this work, we report novel minimal Catcher/Tag pairs engineered from two CnaB domains of a probiotic bacterium, *Lactobacillus plantarum.* We solve the crystal structure of one of the CnaB domains and discuss the design principles for split proteins. The most promising pair, which we here named SilkCatcher/Tag, can catalyze isopeptide-bond mediated protein ligation in high yield, the reaction efficiency being strongly pH dependent. The newly designed pair is not cross-reactive and therefore forms a compatible pair with the SpyCatcher/Tag pair. Finally, we apply the new Catcher/Tag pair together with the SpyCatcher/Tag pair for multiple fragment ligation to produce native-sized spider-silk-like proteins.

Results and Discussion

Engineering novel Catcher/Tag pairs

A cell surface protein lp2578 from a lactic acid bacterium *L. plantarum* was predicted to contain a collagen binding domain and two CnaB domains (Figure 1A, Figure S1A).[28] Both CnaB domains had residues suitable for isopeptide-bond formation, suggesting they would contain intramolecular isopeptide bonds (Figure 1B, Figure S1A). We produced the two CnaB domains, named CnaB1 and CnaB2, recombinantly in *Escherichia coli* as fusion proteins with N-terminally H_6-tagged Smt3-tags, which were subsequently cleaved off with Ulp1 (Figure S1B). Both CnaB domains were expressed in high yields and were soluble at high concentrations. The Far-UV CD spectra of the two proteins were characteristic to CnaB domains, with a positive peak at $\approx$232 nm and a negative peak at $\approx$216 nm (Figure S1C). Furthermore, the molecular weights of the CnaB1 and CnaB2 domains obtained by mass spectrometry corresponded to the sizes expected for a product containing one isopeptide bond each (Figure S1D and S1E).

While mass spectrometry data and sequence homology strongly indicated that the two proteins would contain isopeptide bonds, we wanted to obtain more detailed structural information to aid the design of a Catcher/Tag pair. We solved the crystal structure of the CnaB2 domain with 1.86 Å resolution (Figure 1C and D, Supporting Information Table S1). CnaB2 domain had an unusual crystal packing, forming a hollow sphere, each ball consisting of 8 trimers and altogether 24 domains, with 12 domains within the symmetric unit (data not shown). The CnaB2 domain had typical CnaB-fold consisting of seven β-strands (Figure 1C–F).

The crystal structure verified that the CnaB2 domain contained one isopeptide bond connecting the residues Lys9 in the first β-sheet and Asp91 in the last β-sheet (Figure 1C), the bond being clearly visible in the electron density map. Comparison of the crystal structure to the model at the AlphaFold protein structure database[32,33] shows that although the overall fit of the model is very good, it is missing the isopeptide bond.

We were not able to obtain diffracting crystals from the CnaB1 domain, most likely due to the extremely high solubility of the protein. Instead, a homology model was built with SWISS-MODEL[34] based on the structure of the CnaB2 domain, followed by manual adjustment (Figure 1E). Interestingly, CnaB1 (99aa) and CnaB2 (95aa) domains are smaller than the two CnaB domains used as a starting point for the engineering of the three versions of SpyCatcher/Tag (SpyCatcher/Tag1–3) (121aa), and SnoopCatcher/Tag and DogCatcher/Tag pairs (127aa) (Figure S2).

We engineered Catcher/Tag pairs from CnaB1 and CnaB2 domains by splitting at the loop between the β-strands numbered six and seven (Figure 1D–F). There are no clear guidelines for selecting the split site for protein fragment complementation, and the optimal solution is application-dependent.[35–37] The split site was thus selected based on the design of SpyCatcher/Tag system.[6,7] The Catcher/Tag pairs are called CnaB1Catcher/Tag(S6/7a) and CnaB2Catcher/Tag(S6/7) with the number in parenthesis indicating the location of the split site (Figure 1B). The Catcher/Tag pairs were tested in a model system, in which Catcher was fused to an N-terminal H_6-Smt3 tag (H_6-Smt3-Catcher), whereas Tag was fused to a C-terminal cellulose binding module (CBM) and a H_6-tag (Tag-CBM-H_6) (Figure 1G). Both CnaB1 and CnaB2 Catcher/Tag pairs were able to catalyze protein ligation in optimized conditions as observed by analysis with SDS-PAGE, in which a protein corresponding to the size of a ligation product appeared upon mixing (Figure 1H–J). Instead, no ligation was observed when the isopeptide-forming Asp in CnaB2Tag was mutated to Ala (Figure S3A). The ligation yield of the CnaB1Catcher/Tag(S6/7a) pair was comparable to that of SpyCatcher/Tag (Figure 1J),[6] which prompted us to study the newly engineered CnaB1Catcher/Tag pair in more detail.

Analysis of the ligation product of CnaB1Catcher/Tag-(S6/7a) pair by mass spectrometry confirmed the presence of an isopeptide bond connecting the Catcher/Tag pair (Figure 2A). To determine which one of the two Asp residues located near the predicted isopeptide bond (Asp 92 and 93) was participating in the reaction, we tested ligation reaction after mutating them into alanine one-by-one (Figure 1B, Figure 2B). The mutant harboring Asp93Ala mutation was inactive, whereas the Asp92Ala mutant was active, showing that Asp93 was forming the isopeptide bond. Finally, the association and folding of the CnaB1Catcher and CnaB1Tag could be followed by CD-spectroscopy, in which the characteristic negative peak around 216 nm and positive peak around 232 nm appeared after incubating the Catcher/Tag pair together (Figure 2C).

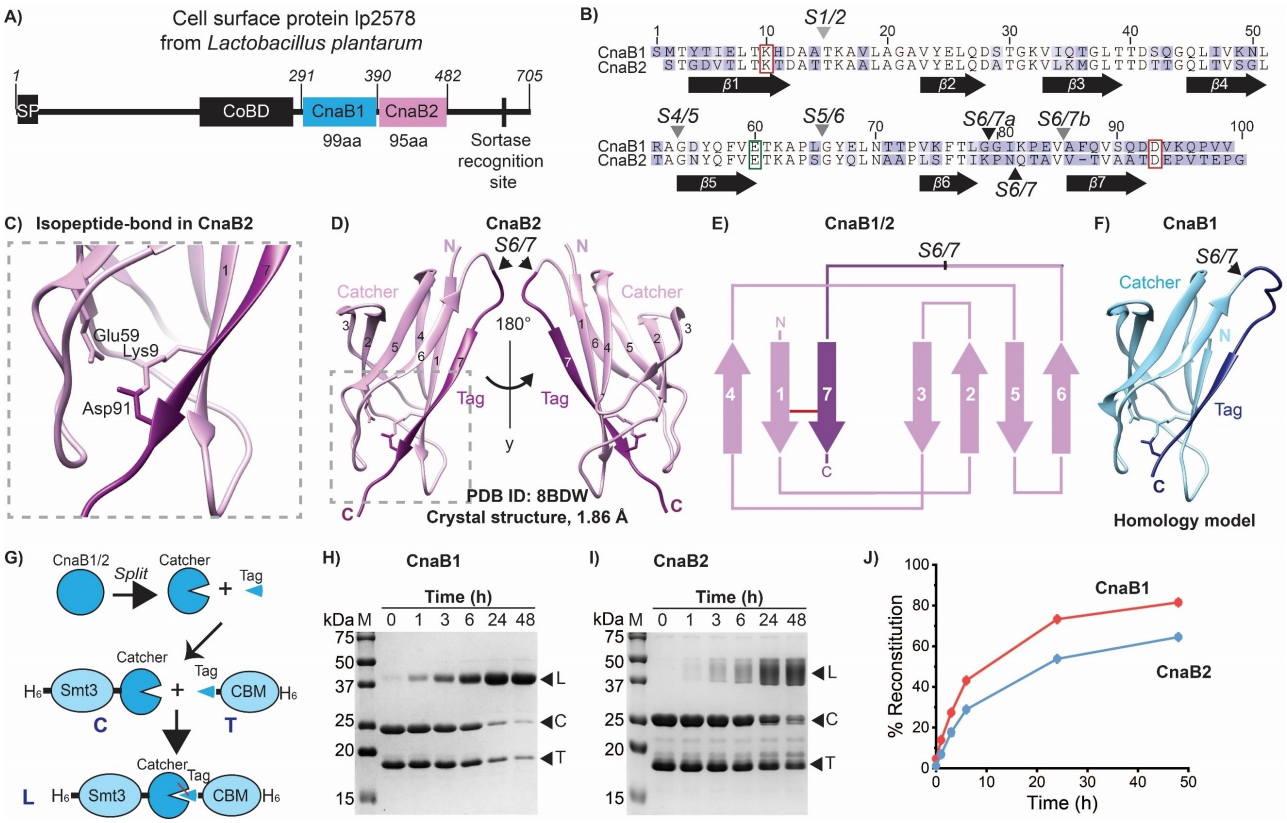

Figure 1. Engineering Catcher/Tag pairs from the *L. plantarum* CnaB1 and CnaB2 domains. A) Schematic presentation of the domain structure of *L. plantarum* cell surface protein lp_2578. Key amino-acid sequence-numbers are shown. SP stands for signal peptide, CoBD stands for collagen binding domain. B) Sequence alignment of CnaB1 and CnaB2 domains. Lys and Asp participating in the reaction are highlighted with red boxes and the catalytic Glu with a green box. Split sites (S) are numbered according to the flanking β-sheets (preceding/following) and indicated with arrowheads. C) Isopeptide bond in the newly solved *L. plantarum* CnaB2 structure (PDB ID: 8BDW). D) Cartoon presentation of the crystal structure of *L. plantarum* CnaB2 domain. Tag (dark purple) and Catcher (light purple) regions of the first tested split variants are shown. β-sheet numbers are given and the side chains of the active site residues are shown with stick model. Split sites are shown with arrow heads. E) Topology diagram of *L. plantarum* CnaB1 and 2 domains based on the homology model and crystal structure. Tag (dark purple) and Catcher (light purple) regions of the first tested split variants are shown. F) Cartoon presentation of the homology model of *L. plantarum* CnaB1 domain. Tag (dark blue) and Catcher (light blue) regions of the first tested split variants are shown. The side chains of the active site residues are shown with a stick model. G) Schematic presentation of engineering split Catcher/Tag pairs from CnaB domains and their fusion to H_6-Smt3 and CBM-H_6 tags for purification and ligation studies. H), I) Analysis of the time scale of the ligation of CnaB1(S6/7a) and CnaB2(S6/7) Catcher/Tag pairs on SDS-PAGE. Smearing of the band corresponding to the ligation product is occasionally observed in CnaB domains and Catcher/Tag pairs[29,30] M stands for molecular weight marker. L, C, and T stand for ligation product (H_6-Smt3-Catcher-Tag-CBM-H_6), Catcher-precursor (H_6-Smt3-Catcher) and Tag-precursor (Tag-CBM-H_6), respectively. J) Reaction rate of the CnaB1 and CnaB2 S6/7 Catcher/Tag pairs. Molecular graphics were prepared with UCSF Chimera.[31]

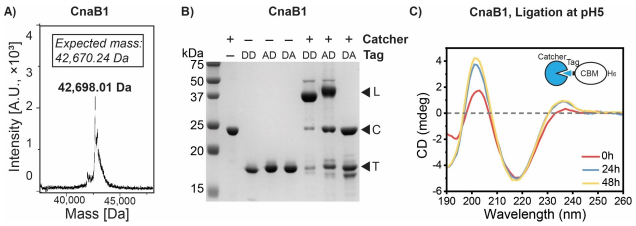

Figure 2. CnaB1Catcher/Tag(S6/7a) ligation. A) Molar mass of the ligation product analyzed by mass spectrometry. B) SDS-PAGE gel showing the ligation reaction of the wildtype Catcher with the wildtype Tag (Asp92 and Asp93; DD), and the two alanine mutants; Asp92Ala (AD) and Asp93Ala (DA) Tag. L, C, and T, stand for ligation product, Catcher, and Tag. C) CD spectra from ligation reaction of CnaB1Catcher/Tag pair after 0 h, 24 h, and 48 h.

Optimization of the ligation reaction

We next studied the optimal reaction conditions. Reaction of Tag-precursor could be driven in close to 100 % when mixed with an excess of the Catcher part (Figure 3A). Reaction was found to work in all tested temperatures 4–45 °C, while being optimal at 37 °C (Figure 3B). The ligation yield of the CnaB2Catcher/Tag pair decreased after cleaving of the solubility tag (Figure S3B–D), while that of the CnaB1 Catcher/Tag pair was not affected (Figure S4A).

The reaction had a pH optimum at pH 5, in which ≈80 % of the precursors had reacted, whereas at pH 3 and pH 7 the yield was just 30 % and at pH 8 and pH 9 only very faint bands for the ligation product were seen (Figure 3C). The ligation reaction of the SpyCatcher/Tag variants 1–3

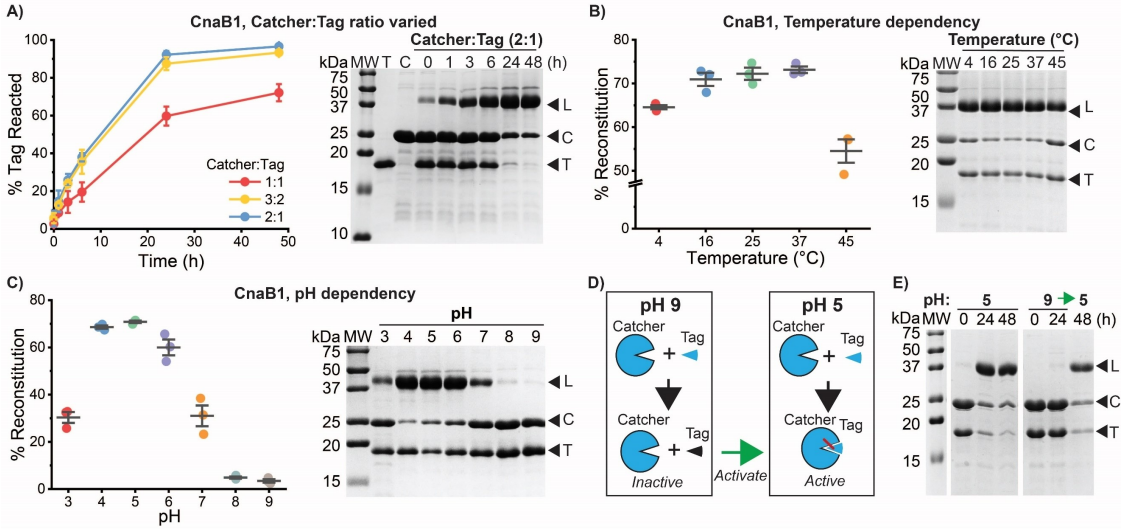

Figure 3. Optimization of CnaB1Catcher/Tag(S6/7a) ligation. A) Adding excess of one of the precursors can drive reaction to completion. B), C) Temperature and pH dependency of the reaction. D), E) Schematic presentation and SDS-PAGE analysis showing recovery of the ligation reaction upon changing pH. M stands for molecular weight marker. L, C, and T stand for ligation product (H_6-Smt3-Catcher-Tag-CBM-H_6), Catcher-precursor (H_6-Smt3-Catcher) and Tag-precursor (Tag-CBM-H_6), respectively. Data show the mean values ± std of triplicate experiments.

and another Catcher/Tag pair, called SdyCatcher/Tag,[16] are also optimal at pH 5–6.[6,11,26] All these pairs have Lys and Asp as the isopeptide bond forming residues. In contrast, SnoopCatcher and DogCatcher, in which an isopeptide bond forms between Lys and Asn, have optimal reactions at pH > 9.[14,25] CD spectra of the full-length CnaB1 domain and the CnaB1Catcher(N6/7a) show that the folding of the domains is not pH-sensitive (Figure S4B and S4C). The optimal pH range is most likely dependent on the pK_a of the isopeptide-bond forming residues, which is affected by the local environment.[30,38] The detailed mechanism behind this observation remains to be determined in future studies.

The similar optimal pH to that of SpyCatcher/Tag1–3 provides possibilities for their use in single-pot reactions. Inspired by the drastic differences in activity depending on pH, we wanted to test the possibility to induce the reaction by a pH switch. Indeed, almost no reaction took place after 24 h incubation at pH 9, but the activity was retained after lowering pH to 5 (Figure 3D and E).

To study ligation activity of the newly designed pair in vivo, the two precursors coding for the Catcher and Tag were subcloned into vectors carrying different origins of replications and antibiotic resistances for coexpression.[39] (Figure 4A). Ligation product was observed despite the pH ≈7.4–7.8 of *E. coli* cytoplasm[40] being non-optimal (Figure 4B). Adjusting the expression levels of the two precursors to overexpress the Tag-precursor without an H_6-tag drove the ligation reaction of the Catcher-precursor into near completion allowing us to purify only the ligation product (Figure 4B).

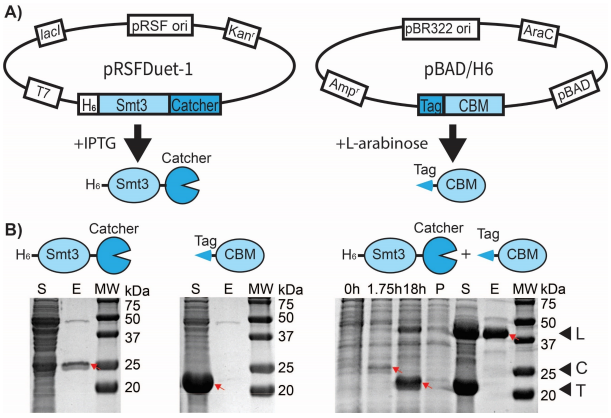

Figure 4. In vivo ligation with CnaB1Catcher/Tag(S6/7a). A) Schematic presentation of the plasmid construction. B) SDS-PAGE gels showing expression of the two precursors alone and as coexpression. Timescale of the expression is shown for the coexpression. S stands for supernatant, P for pellet, E for elution, and MW for molecular weight marker. The ligation product (L), Catcher (C), and Tag (T) are depicted with red arrows.

Analysis of split variants

We wanted to explore how the split site affects ligation yield and compared four Catcher/Tag variants, split at four different loops (Figure 5). We named the split sites according to the β-strands flanking the loop that was split, and call them S1/2, S4/5, S5/6, and S6/7. The longer one of the two split fragments is always called Catcher and the shorter one Tag. Catcher/Tag pairs are typically engineered by splitting CnaB domain either at the loop corresponding to that of the split site S1/2 (SnoopCatcher/Tag)[14] or that of split site 6/7 (SpyCatcherTag[6]) in Figure 5A. In addition to systematically testing variants split at four different loops, we

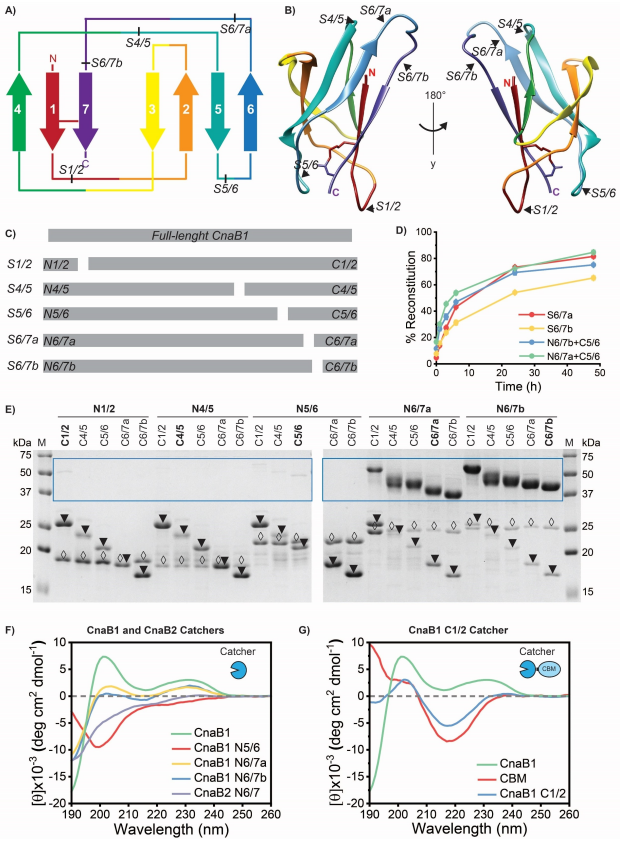

Figure 5. Split sites in CnaB1. A), B) Different split sites shown in the topology map (A) and cartoon presentation (B) of the homology model of CnaB1. C) Schematic presentation of the lengths and the naming of the split fragments. D) Plot showing the reaction kinetics of selected pairs. Data show the mean values ± std of triplicate experiments. Some of the error bars are too small to be visible. E) SDS-PAGE analysis of the cross-reactivity of the different split fragments. M stands for molecular weight marker. The area in which the ligation products appear is highlighted with a blue box, N- and C-precursor are depicted with open diamonds and filled triangles, respectively. F), G) CD spectra of selected CnaB1 and CnaB2 Catchers. Full-length CnaB1 is shown as a reference.

included a variant of the original S6/7a split site in the comparison. This variant (S6/7b) was split five residues closer to the C-terminus. In addition, S6/7b Tag (N6/7b) is three residues shorter from the C-terminus. While minor ligation activity was observed for all the split variants, only the two Catcher/Tag pairs split at the last loop (S6/7) were found to have high yields >60 % (Figure 5C–E). The yield from the S6/7b split variant was worse than that of the S6/7a (Figure 5D), despite the difference in split site is just 5 aa, most likely due to the observed decrease in stability. Interestingly, mixing the S5/6 Tag, which was not able to react with the S5/6 Catcher, with either of the Catchers split at the S6/7 site, improved the ligation yield (Figure 5D). Studying cross-reactivity of the different split variants showed that while only short gaps can be tolerated, major overlap of the fragments does not prevent the reaction (Figure 5E). The Catcher/Tag pairs split at the S6/7 loop did cross-react with each other and the three residues missing did not prevent the reaction. However, any larger gap between the Catcher and Tag fully abolished the reaction. Instead, major overlap of the Catcher and Tag fragments did not prevent the ligation reaction. The two Catchers split at the S6/7 site were able to mediate ligation with all the tested Tags. Even the S1/2 Catcher and the S6/7b Catcher, which have a 71aa overlap, retained ligation activity together. Similar results were reported for the *E. coli* glycinamide ribonucleotide formyltransferase and, as an extreme example, in the domain swapping of inteins.[41,42]

To better understand the differences in ligation activities, we studied the folding of the Catcher variants using Far-UV CD spectroscopy. CnaB domain has a characteristic CD spectrum with two positive peaks at ≈200 nm and at ≈230 nm (Figure 5F). The CD spectra of the two functional CnaB1 Catchers split at the last loop (S6/7) indicated that they were partially folded, showing smaller positive peaks at both ≈200 nm and ≈230 nm (Figure 5F). In contrast, significantly less folded structures were observed for the Catcher variants split at the S5/6 and at the S1/2 loops (Figure 5F and G). The Catcher variant split at the S4/5 loop could not be analyzed, because it did not remain soluble after cleaving the solubility tag (Figure S4E). In addition, the CnaB2 Catcher split at the same S6/7 loop was mainly in a random coil conformation (Figure 5F). The observed differences in the folding of CnaB1 and CnaB2 (S6/7) Catchers were not dependent on the pH (Figure S4C and S4D). Splitting deeper into the protein often promotes association but may also lead to misfolding or prevent folding due to the exposure of hydrophobic regions.[36] This could also explain the poor solubility of the N4/5 Catcher. We interpret that despite having some degree of structure, the two Catchers are not in a fully folded state, because a significant increase in the intensity of the two positive peaks is observed to take place along the ligation reaction (Figure 2C). The data indicates that partial folding, perhaps into a molten globule type structure, improves protein fragment complementation and thus also ligation yield. It is known that sites essential for folding are typically not suitable for splitting,[34] which can explain why no ligation was observed when split at the loop closest to the N-terminus (S1/2). In addition to the high ligation yield, the solubility of the CnaB1N6/7a Catcher was an important factor for selecting it to further studies and could be essential also for other biotechnological applications.

Ligation of spider silk proteins

Large-scale production of spider silk proteins has been hindered by difficulties in the recombinant production of native-sized spider silk proteins in high yields and in a soluble form.[18] Attempts to express long dragline silk sequences in production organisms have generally resulted in truncated proteins, low expression yields, and/or insoluble proteins, although the fibers spun from them have had promising properties.[21,43,44] Therefore, we employed an alternative strategy, in which shorter but soluble spidroins are covalently conjugated in vitro using isopeptide-bond

formation as initially demonstrated with the SpyCatcher/Tag protein-peptide pair. We have previously shown that SpyCatcher/Tag could be used to ligate 2–3 of 43 kDa fragments of the dragline silk repeat sequence ADF3 from *Araneus diadematus* (ADF3).[45,46] The length of the silk fragments was, however, limited by the lack of Catcher/Tag pairs that could be used together with SpyCatcher/Tag pair for multiple fragment ligation. Here, we wanted to test, if the newly engineered Catcher/Tag pair (CnaB1 S6/7) would allow us to overcome this challenge. We decided to call the new pair SilkCatcher/Tag pair.

Two prerequisites for the success of the strategy are that the SilkCatcher/Tag pair should not be cross-reactive with the SpyCatcher/Tag-pair and it should be possible to produce in high yields and in a soluble form as a fusion protein with silk. We first studied cross-reactivity with model proteins and no cross-reactivity between the SpyCatcher/Tag and SilkCatcher/Tag pairs was observed (Figure S5). Next, we constructed fusion proteins of the ADF3 repeat sequence[46] with the SilkCatcher and a cellulose binding module (CBM) from the *Clostridium thermocellum* cellulosome.[47] Fusion with CBM improves solubility of the silk protein and allows utilizing it as a composite material with cellulose as well as an adhesive on delignified cellulose.[48,49] The fusion protein could be expressed in high yields and purified by heat precipitation at 70 °C (Figure S6). Attempt to use SnoopCatcher as a reference showed that the combination with ADF3 was not soluble and was therefore not pursued further.

We designed two sets of three precursors (Figure 6A) that allowed us to covalently ligate 4× and 5× silk proteins using a combination of the SpyCatcher/Tag and SilkCatcher/Tag pairs. Five-fragment conjugation could be achieved with only three precursors because Catcher and Tag can be fused in both the N- and C-termini of the target protein.[22] The pH optimum of SilkCatcher is very close to that of SpyCatcher, making it easy to find suitable conditions for one-pot ligation (Figure 3 and Ref. [11]). Indeed, we were able to ligate 4× and 5× silk proteins in a 48 h one-pot ligation (Figure 6B and C). The yield from especially the 5× silk protein ligation reaction was, however, not satisfactory. Attempts to optimize the ligation conditions further did not improve the yield. We have observed that the ligation efficiency of Catcher/Tag pairs may vary depending on the other parts of fusion proteins. Fusion proteins with silk were more prone for aggregation than those with the model proteins and thus more sensitive to the ligation conditions, which can explain the lower than theoretically expected yield.

To overcome the limitations with yield and to minimize the required purification steps, we decided to apply an alternative approach based on the following three concepts: 1) the possibility to drive the reaction yield close to 100 % by adjusting the reaction ratios (Figure 3A), 2) an observation that SilkCatcher/Tag was active in cell lysate, and 3) the design of a set of constructs with and without H_6-tags (Figure 7). Ligation is performed in two steps: first in cell lysate with the SilkCatcher/Tag pair, followed by IMAC purification and subsequent ligation using the SpyCatcher2/Tag pair (Figure 7A). In step 1, all precursors 1–4 were expressed in *E. coli.* After harvesting the cells, cell pellets of H_6-tagged precursor 2 and non-tagged precursors 1 and 3 were mixed in 1:2 ratio, respectively, so that the precursors with H_6-tags were consumed in the ligation reaction. Cell pellet mixtures were lysed and incubated at 37 °C for 48 h prior to purification of the ligation product by IMAC. Because almost no precursors with H_6-tag were left, only the ligation product was obtained in the elution fractions (Figure 7B). In the second step, ligation products from step 1 were mixed in vitro either with each other or with precursor 4 to result in 4× and 5× silks, respectively. Careful adjustment of the precursor ratios and ligation conditions resulted in the reaction proceeding into near completion during a 1 h incubation, with 93 ± 1 % ligation yield observed for both 4× and 5× silks (Figure 7C).

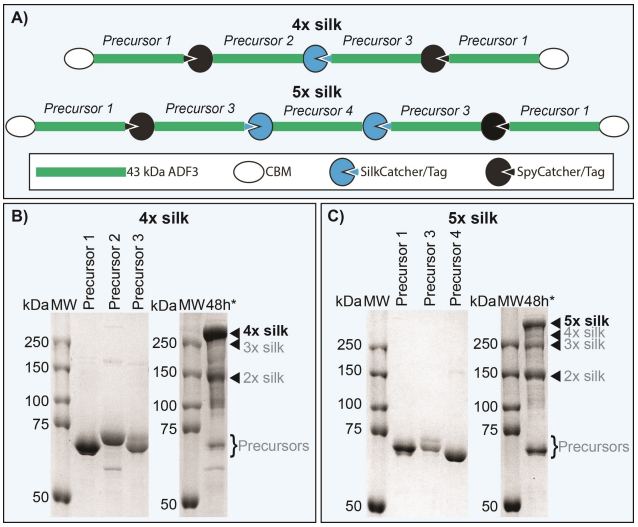

Figure 6. In vitro ligation of 4× and 5× silk. A) Schematic presentation of the constructs and the ligation strategy. B), C) SDS-PAGE gels showing ligation of 4× and 5× silks, respectively.

Conclusion

We report here the design of novel Catcher/Tag pairs for biomolecular click reactions. One of the newly engineered pairs, called SilkCatcher/Tag pair, showed efficient reconstitution yield both under mild in vitro conditions and in vivo. The possibility to regulate the ligation activity by adjusting pH opens new possibilities for the design of pH-inducible modifications. Further optimization of the SilkCatcher/Tag pair may still widen the range of possible applications.

Furthermore, the newly designed SilkCatcher/Tag pair was used to covalently ligate native-sized spider-silk proteins consisting of 4 and 5 of 43 kDa fragments of spider-silk repeat sequence, thus solving one of the key challenges in the production of recombinant spider silk. We were able to achieve >90 % ligation yield by applying a sequential ligation approach. The conjugation of silk proteins using the

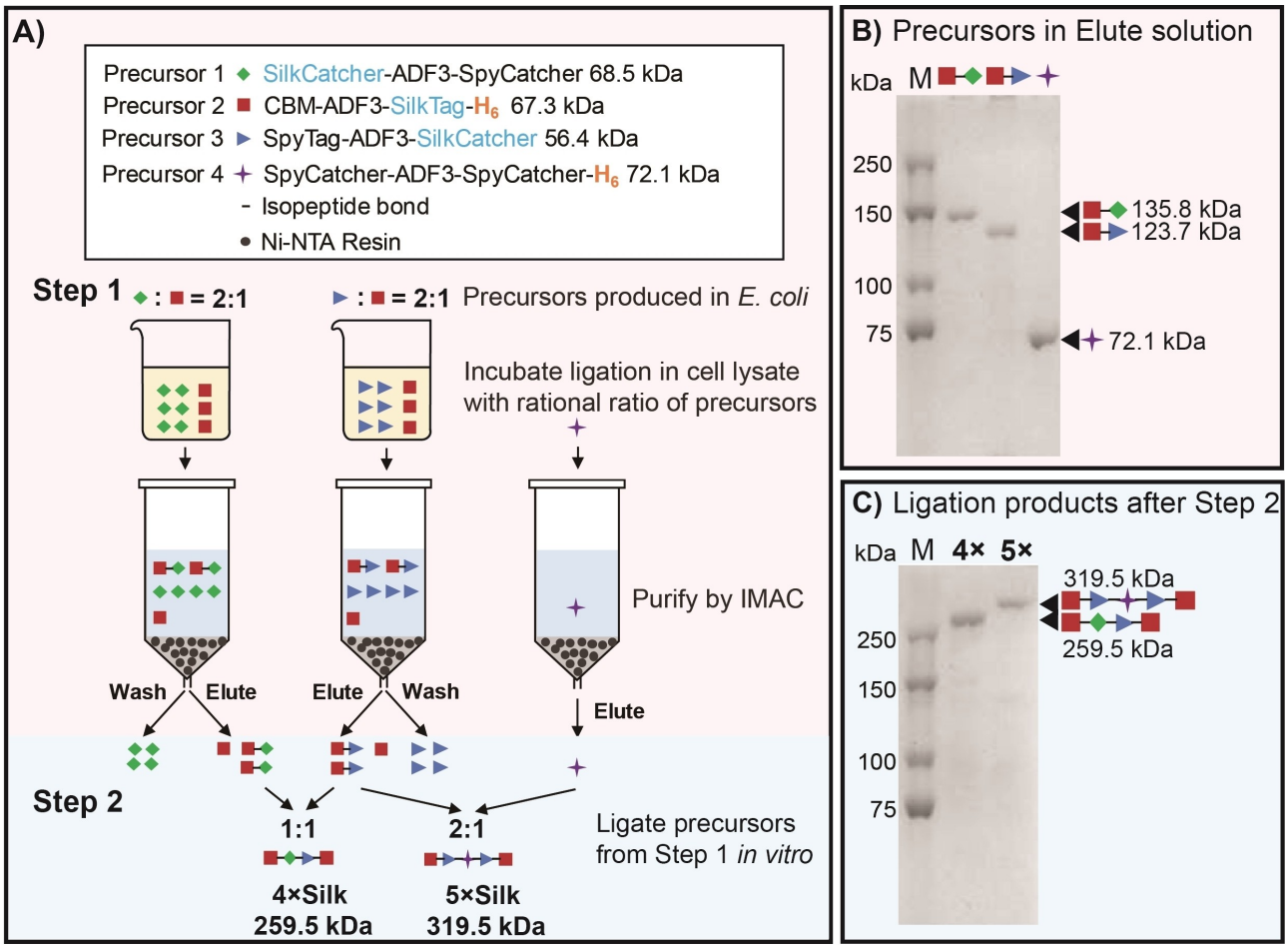

Figure 7. Stepwise ligation of 4× and 5× silk. A) Schematic presentation of the constructs and ligation strategy for 4× and 5× silk. B) An SDS-PAGE gel showing the IMAC purified ligation products after step 1. C) SDS-PAGE gel showing ligation products of step 2 after 1 h incubation. M stands for protein weight marker.

SpyCatcher/Tag protein/peptide pair has significant advantages over previously reported polymerization through dimerization via disulfide bridges or by the protein *trans*-splicing of inteins.[44,50–55] Both the disulfide-bridge and intein-mediated ligation are dependent on the redox state of the system, which limits their applications. In addition, all previously reported cases of the production of the >50 kDa silk proteins were either performed using eggcase, flagelliform, or aciniform silks, which are more soluble but do not have comparable mechanical properties to dragline silks, or resulted in the production of insoluble fusion proteins.

In summary, the newly designed SilkCatcher/Tag pair is efficient in protein ligation, robust, and orthogonal with the widely used SpyCatcher/Tag pair. Therefore, it is a highly valuable addition to the toolbox of protein ligases that can be applied to multiple-fragment ligation approach either together with the SpyCatcher/tag pair or in combination with intein technologies.

Experimental Section

Please see the Supporting Information for details.

Acknowledgements

We thank Eva Crosas for her help with the early phases of the structure refinement. Protein crystallization was performed at SPC facility at EMBL Hamburg and the CD spectroscopy and mass spectrometry of the crystallized proteins at the Center for Structural Systems Biology (CSSB, Deutsches Elektronen-Synchrotron DESY). We acknowledge technical support by the SPC facility at EMBL Hamburg. The synchrotron data was collected at beamline operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). This work was supported by the Academy of Finland through its Centres of Excellence Programme Life-Inspired Hybrid Materials (LIBER, 2022–2029) under project no 346105 and Academy of Finland projects nos. 317395, 308772, and 333238. We are grateful for the support by the FinnCERES Materials Bioeconomy Ecosystem and use of the Bioeconomy Infrastructure at the Aalto University.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: Catcher/Tag · Protein Engineering · Protein Ligation · Protein Modifications · Spider Silk

[1] D. W. Wood, J. A. Camarero, *J. Biol. Chem.* **2014**, *289*, 14512–14519.
[2] B. Zakeri, *ChemBioChem* **2015**, *16*, 2277–2282.
[3] C. Freund, D. Schwarzer, *ChemBioChem* **2021**, *22*, 1347–1356.
[4] A. H. Keeble, M. Howarth, *Chem. Sci.* **2020**, *11*, 7281–7291.
[5] I. N. A. Khairil Anuar, A. Banerjee, A. H. Keeble, A. Carella, G. I. Nikov, M. Howarth, *Nat. Commun.* **2019**, *10*, 1734.
[6] 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*, E690–E697.
[7] G. Veggiani, B. Zakeri, M. Howarth, *Trends Biotechnol.* **2014**, *32*, 506–512.
[8] B. Zakeri, M. Howarth, *J. Am. Chem. Soc.* **2010**, *132*, 4526–4527.
[9] H. J. Kang, F. Coulibaly, F. Clow, T. Proft, E. N. Baker, *Science* **2007**, *318*, 1625.
[10] U. Sridharan, K. Ponnuraj, *Biophys. Rev. Lett.* **2016**, *8*, 75–83.
[11] A. H. Keeble, A. Banerjee, M. P. Ferla, S. C. Reddington, I. N. A. Khairil Anuar, M. Howarth, *Angew. Chem. Int. Ed.* **2017**, *56*, 16521–16525; *Angew. Chem.* **2017**, *129*, 16748–16752.
[12] A. H. Keeble, P. Turkki, S. Stokes, I. N. A. K. Anuar, R. Rahikainen, V. P. Hytönen, M. Howarth, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 26523–26533.
[13] D. Hatlem, T. Trunk, D. Linke, J. C. Leo, *Int. J. Mol. Sci.* **2019**, *20*, 2129.
[14] G. Veggiani, T. Nakamura, M. D. Brenner, R. V. Gayet, J. Yan, C. V. Robinson, M. Howarth, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1202–7.
[15] A. E. L. Busche, A. S. Aranko, M. Talebzadeh-Farooji, F. Bernhard, V. Dötsch, H. Iwaï, *Angew. Chem. Int. Ed.* **2009**, *48*, 6128–6131; *Angew. Chem.* **2009**, *121*, 6244–6247.
[16] J. Shi, T. W. Muir, *J. Am. Chem. Soc.* **2005**, *127*, 6198–6206.
[17] J. A. Kluge, O. Rabotyagova, G. G. Leisk, D. L. Kaplan, *Trends Biotechnol.* **2008**, *26*, 244–251.
[18] M. Ramezaniaghdam, N. D. Nahdi, R. Reski, *Front. Bioeng. Biotechnol.* **2022**, *10*, 835637.
[19] J. L. Yarger, B. R. Cherry, A. Van Der Vaart, *Nat. Rev. Mater.* **2018**, *3*, 18008.
[20] A. Sponner, B. Schlott, F. Vollrath, E. Unger, F. Grosse, K. Weisshart, *Biochemistry* **2005**, *44*, 4727–4736.
[21] O. Tokareva, V. A. Michalczechen-Lacerda, E. L. Rech, D. L. Kaplan, *Microb. Biotechnol.* **2013**, *6*, 651–63.
[22] W. Bin Zhang, F. Sun, D. A. Tirrell, F. H. Arnold, *J. Am. Chem. Soc.* **2013**, *135*, 13988–13997.
[23] 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.
[24] C. Schoene, J. O. Fierer, S. P. Bennett, M. Howarth, *Angew. Chem. Int. Ed.* **2014**, *53*, 6101–6104; *Angew. Chem.* **2014**, *126*, 6215–6218.
[25] A. H. Keeble, V. K. Yadav, M. P. Ferla, C. C. Bauer, E. Chuntharpursat-Bon, J. Huang, R. S. Bon, M. Howarth, *Cell Chem. Biol.* **2022**, *29*, 339–350.
[26] L. L. Tan, S. S. Hoon, F. T. Wong, *PLoS One* **2016**, *11*, e0165074.
[27] M. Pröschel, M. E. Kraner, A. H. C. Horn, L. Schäfer, U. Sonnewald, H. Sticht, *PLoS One* **2017**, *12*, e0179740.
[28] J. Boekhorst, M. Wels, M. Kleeberezem, R. J. Siezen, *Microbiology* **2006**, *152*, 3175–3183.
[29] L. Li, J. O. Fierer, T. A. Rapoport, M. Howarth, *J. Mol. Biol.* **2014**, *426*, 309–317.
[30] R. M. Hagan, R. Björnsson, S. A. McMahon, B. Schomburg, V. Braithwaite, M. Bühl, J. H. Naismith, U. Schwarz-Linek, *Angew. Chem. Int. Ed.* **2010**, *49*, 8421–8425; *Angew. Chem.* **2010**, *122*, 8599–8603.
[31] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
[32] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
[33] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, *Nucleic Acids Res.* **2022**, *50*, D439–D444.
[34] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, *Nucleic Acids Res.* **2018**, *46*, W296–W303.
[35] S. S. Shekhawat, I. Ghosh, *Curr. Opin. Chem. Biol.* **2011**, *15*, 789–797.
[36] O. Dagliyan, A. Krokhotin, I. Ozkan-Dagliyan, A. Deiters, C. J. Der, K. M. Hahn, N. V. Dokholyan, *Nat. Commun.* **2018**, *9*, 4042.
[37] T. B. Dolberg, A. T. Meger, J. D. Boucher, W. K. Corcoran, E. E. Schauer, A. N. Prybutok, S. Raman, J. N. Leonard, *Nat. Chem. Biol.* **2021**, *17*, 531–539.
[38] X. Hu, H. Hu, J. A. Melvin, K. W. Clancy, D. G. McCafferty, W. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 478–485.
[39] S. Züger, H. Iwai, *Nat. Biotechnol.* **2005**, *23*, 736–740.
[40] J. L. Slonczewski, B. P. Rosen, J. R. Alger, R. M. Macnab, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 6271–6275.
[41] M. Ostermeier, A. E. Nixon, J. H. Shim, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3562–3567.
[42] A. S. Aranko, J. S. Oeemig, T. Kajander, H. Iwaï, *Nat. Chem. Biol.* **2013**, *9*, 616–622.
[43] X.-X. Xia, Z.-G. Qian, C. S. Ki, Y. H. Park, D. L. Kaplan, S. Y. Lee, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14059–14063.
[44] S. Lin, G. Chen, X. Liu, Q. Meng, *Biopolymers* **2016**, *105*, 385–392.
[45] L. Lemetti, A. Scacchi, Y. Yin, M. Shen, M. B. Linder, M. Sammalkorpi, A. S. Aranko, *Biomacromolecules* **2022**, *23*, 3142–3153.
[46] J. M. Gosline, P. a Guerette, C. S. Ortlepp, K. N. Savage, *J. Exp. Biol.* **1999**, *202*, 3295–3303.
[47] J. Tormo, R. Lamed, J. Chirino, E. Morag, E. Bayer, Y. Shoham, T. Steitz, *EMBO J.* **1996**, *15*, 5739–5751.

[48] P. Mohammadi, A. S. Aranko, C. P. Landowski, O. Ikkala, K. Jaudzems, W. Wagermaier, M. B. Linder, *Sci. Adv.* **2019**, *5*, eaaw2541.
[49] L. Lemetti, J. Tersteegen, J. Sammaljärvi, A. S. Aranko, M. B. Linder, *ACS Sustainable Chem. Eng.* **2022**, *10*, 552–561.
[50] S. Grip, J. Johansson, M. Hedhammar, *Protein Sci.* **2009**, *18*, 1012–1022.
[51] Z. Lin, Q. Deng, X. Y. Liu, D. Yang, *Adv. Mater.* **2013**, *25*, 1216–1220.
[52] L. Xu, J. K. Rainey, Q. Meng, X.-Q. Liu, *PLoS One* **2012**, *7*, e50227.
[53] V. Hauptmann, N. Weichert, M. Menzel, D. Knoch, N. Paege, J. Scheller, U. Spohn, U. Conrad, M. Gils, *Transgenic Res.* **2013**, *22*, 369–377.
[54] M.-L. Tremblay, L. Xu, T. Lefèvre, M. Sarker, K. E. Orrell, J. Leclerc, Q. Meng, M. Pézolet, M. Auger, X.-Q. Liu, J. K. Rainey, *Sci. Rep.* **2015**, *5*, 11502.
[55] C. H. Bowen, B. Dai, C. J. Sargent, W. Bai, P. Ladiwala, H. Feng, W. Huang, D. L. Kaplan, J. M. Galazka, F. Zhang, *Biomacromolecules* **2018**, *19*, 3853–3860.

Manuscript received: November 7, 2022
Accepted manuscript online: January 25, 2023
Version of record online: February 7, 2023