FEBS Letters

FEBS PRESS science publishing by scientists

Segmental isotopic labeling of a single-domain globular protein without any refolding step by an asparaginyl endopeptidase

Kornelia M. Mikula, Igor Tascón*, Jenni J. Tommila and Hideo Iwaï

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Finland

Correspondence
H. Iwaï, Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki. P.O. Box 65, Helsinki, FIN-00014, Finland
Tel: +358 9 191 59752
E-mail: hideo.iwai@helsinki.fi

***Present address**
Institute of Biochemistry, Goethe Frankfurt University, Frankfurt/Main, Germany

(Received 27 February 2017, revised 23 March 2017, accepted 24 March 2017, available online 20 April 2017)

doi:10.1002/1873-3468.12640

Edited by Stuart Ferguson

Asparaginyl endopeptidases (AEPs) catalyze head-to-tail backbone cyclization of naturally occurring cyclic peptides such as cyclotides, and have become an important peptide-engineering tool for macrocyclization and peptide ligation. Here, we report efficient protein ligation in *trans* by mimicking efficient backbone cyclization by an AEP without any excess of reactants. We demonstrate a practical application of segmental isotopic labeling for NMR studies of a single-domain globular protein without any refolding step using the recombinant AEP prepared from *Escherichia coli*. This simple protein ligation approach using an AEP could be applied for incorporation of various biophysical probes into proteins as well as post-translational production of full-length proteins.

Keywords: asparaginyl endopeptidase; NMR; protein ligation; segmental isotopic labeling

Selective ligation of two or more polypeptides with peptide bonds has increasingly become important for incorporating various biochemical and biophysical probes such as stable-isotopes, fluorophores, drugs, and unnatural amino acids into proteins. Ligation of protein fragments with peptide bonds could also facilitate production of toxic proteins *in vitro*. Protein ligation has thus opened various peptide and protein engineering possibilities such as head-to-tail backbone cyclization. Ligation of a synthetic unprotected peptide or a recombinant protein/peptide with α-thio-ester and a peptide bearing an N-terminal cysteine has been achieved by native chemical ligation (NCL; Fig. 1A) [1]. This reaction requires a Cys residue at the junction and is limited by production of thio-ester modified proteins, which can be produced either by chemical synthesis, by thiolysis of recombinant intein-fusion proteins [expressed protein ligation (EPL)], or by enzyme-mediated modification [2–4]. Protein ligation by protein *trans*-splicing (PTS) via split inteins or split HINT domains is not restricted by Cys residue at the ligation junction (Fig. 1B) [5–11]. However, the target protein sequences have to be genetically fused with split intein fragments and the split protein fragments have to be reconstituted into an active conformation for protein ligation. This procedure can be cumbersome and time-consuming [10,11]. The fusion proteins with split intein fragments could also be insoluble,

Abbreviations
AEPs, asparaginyl endopeptidases; DTT, dithiothreitol; EPL, expressed protein ligation; GB1, B1 domain of IgG-binding protein G; GFP, green fluorescent protein; IMAC, immobilized metal ion affinity chromatography; IPTG, isopropyl-β-D-thiogalactoside; MBP, maltose-binding protein; NCL, native chemical ligation; PBS, phosphate-buffered saline; PTS, protein *trans*-splicing; SB, Super Broth; SrtA, *Staphylococcus aureus* sortase A; TCEP, tris(2-carboxyethyl)phosphine.

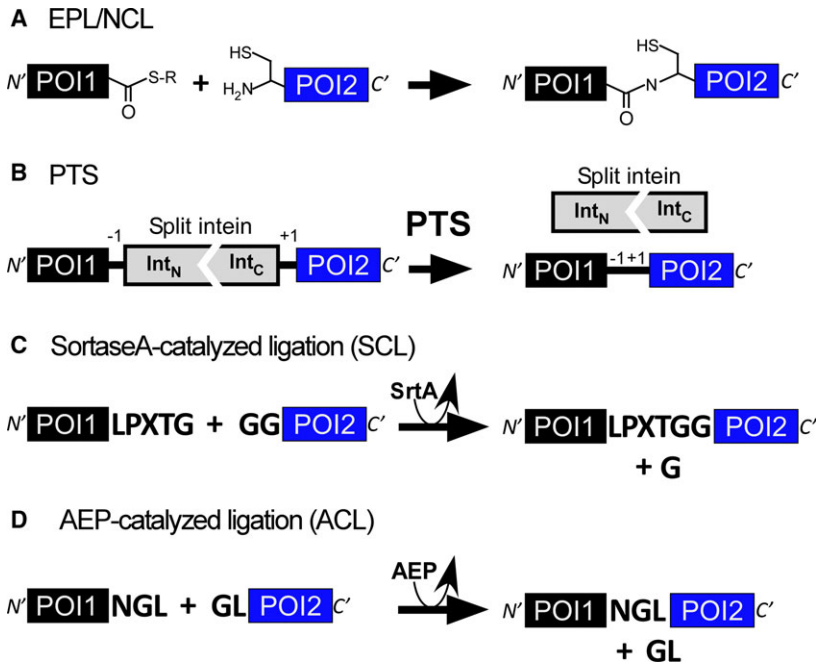

Fig. 1. Commonly used protein ligation methods. (A) NCL and EPL. (B) PTS via split inteins. (C) SrtA-catalyzed protein ligation with the recognition sequence. (D) AEP-catalyzed protein ligation with the reported recognition sequence of *Oa*AEP1$_b$.

thereby requiring additional refolding steps [8–11]. Even though the solubility issue of the precursors fused to split intein fragments has been recently alleviated by highly soluble inteins from halophilic organisms, preparations of individual split intein fusions are mandatory [12].

An alternative approach is enzyme-catalyzed protein/peptide ligation. Formally, peptide ligation is reversal of hydrolysis of a peptide bond by proteases. Thus, proteases have been used for peptide ligation by shifting the equilibrium toward ligation from hydrolysis by changing the solvent conditions using organic solvents, or using one of the reactants in excess, and/or by using the proximity effect [13,14]. This approach has been applied only to a limited number of peptides and proteins [13].

As reversal of protease activities has been very limited, enzymes with *trans*-peptidase activity have been designed or identified in nature. An engineered serine protease of subtilisin called subtiligase has been demonstrated to ligate peptide fragments via thio-ester intermediate [15]. Sortase-catalyzed protein ligation using *Staphylococcus aureus* sortase A (SrtA), which cross-links surface proteins to the peptidoglycan in the cell wall of Gram-positive bacteria, has been widely used for peptide ligation in labeling and cyclization [16,17]. SrtA catalyzes *trans*-peptidation by cleaving between threonine and glycine within 'LPXTG' recognition motif and subsequently joining the carboxyl group of threonine to an amino group of di- or triglycine of an incoming C-terminal peptide (Fig. 1C) [16]. This SrtA-catalyzed protein ligation can be limited by the insertion of 'LPXTG' sequence in the ligated product, although SrtA-catalyzed protein ligation using engineered sortase with altered specificities have been recently developed [17,18]. Due to this sequence insertion, the ligated product remains the substrate of SrtA, which additionally reduces final yields.

Some of naturally occurring cyclic peptides have been produced by head-to-tail backbone cyclization by enzymes such as PatG and asparaginyl endopeptidases (AEPs) [19–22]. In particular, cyclotides such as kalata B1 from plants have been efficiently cyclized by AEPs from plants such as butelase 1 from *Clitoria ternatea* and AEP1$_b$ from *Oldenlandia affinis* (*Oa*AEP1$_b$) [21,22]. As cyclization and protein ligation in *trans* is an identical *trans*-peptidase reaction, enzymes capable of backbone cyclization can be used for bimolecular peptide ligation in *trans* depending on the conditions [21–24]. When thermodynamically controlled, the simplest way to shift the equilibrium toward peptide ligation is to supply one of the reactants in excess. Indeed a fivefold excess of one of the reactants has been used for bimolecular ligation in *trans* by butelase 1 [21,23]. In contrast, backbone cyclization of cyclotides and other small peptides by AEPs is very efficient reaching > 95% backbone cyclization. This suggests that the head-to-tail peptide formation is favorable compared with bimolecular *trans* reaction due to their close proximity [21,22].

Here, we present an efficient protein ligation approach by the proximity effect using an AEP and

demonstrate a practical application in segmental isotopic labeling of a single-domain globular protein without any refolding process.

Materials and methods

Construction of plasmids

Plasmid for production of AEP1 from *Oldenlandia affinis*

The codon optimized gene of AEP1 from *O. affinis* (*Oa*AEP1) (UniProtKB Entry name: A0A0N9JZ32_OLDAF) was synthesized and purchased from Integrated DNA Technologies, Inc. USA. The gene corresponding to the residues 24–474 of *Oa*AEP1 was cloned between *Bam*HI and *Hind*III sites of pHYRSF53, resulting in pBHRSF184 [25]. This plasmid expresses the N-terminally His-tagged SUMO-*Oa*AEP1 fusion protein, instead of UBQ fusion of *Oa*AEP1$_b$ originally reported [22]. This plasmid for *Oa*AEP1 production is deposited at www.addgene.org/Hideo_Iwai with ID: 89482.

Plasmids for backbone cyclization assay

The gene of green fluorescent protein (GFP) was amplified together with the N-terminal hexahistidine tag and the C-terminal thrombin cleavage sequence from pIWT55-63his using two oligonucleotides of I965: 5′-CACAGAGAACAG ATTGGTGGTGGATTACCACATCATCACCACCATCACTC and I964: 5′-AGGAAGCTTACAGTCCATTTCGGGTACTACCGCGTGGCACCAACCCAG [26].

The gene of yeast SMT3 (SUMO) domain without the N-terminal hexahistidine tag was amplified from pHYRSF53 using I916: 5′-ACATATGGGCGACTCAGAAGTCAATC and HK157: 5′-ACATATGGCTCTTCGAACCACCAATCTGTTCTCTGTGAG [25]. The two PCR products were assembled and reamplified using the two oligonucleotides of I916 and I964. The amplified assembled PCR product was cloned into pHYRSF194 using *Nde*I and *Hind*III sites, resulting in pITRSF4 encoding SMT3-GLPH$_6$-(GFP)-LVPRGSTRNGL. pHYRSF194 was derived from pHYRSF1 by mutating *Nco*I site to *Nde*I site at the N terminus [12,27]. Plasmid pJTRSF50 encoding SMT3-GLPH$_6$-(GFP)-LVPRGSTRNCL was constructed by PCR using the two oligonucleotide of I916 and J052: 5′-AGGAAGCTTACAGACAATTTCGGGTACTACCGCG, followed by cloning into pHYRSF194 between *Nde*I and *Hind*III sites. The vector pJTRSF55 encoding SMT3-GLPH$_6$-(GFP)-LVPRGSTRNAL was similarly constructed using the two oligonucleotides of I916 and J083: 5′-AGGAAGCTTACAGAGCATTTCGGGTACTACCGC. The vectors of pITRSF4, pJTRSF50, and pJTRSF55 are deposited at www.addgene.org/Hideo_Iwai with IDs: 89699, 89700, and 89701, respectively.

Plasmids of substrates for bimolecular protein ligation

For protein ligation in *trans*, GFP with the N-terminal 'GLP' sequence was constructed by amplifying the SMT3-GFP fusion from pITRSF4 with the two oligonucleotides of I916 and #153: 5′-CGCAAGCTTAAGTCAACCCAGCAGCAG and cloned into pHYRSF194 using *Nde*I and *Hind*III sites, resulting in pBHRSF208. This vector encodes SMT3-GLPH$_6$-(GFP). The gene of B1 domain of IgG-binding protein G (GB1) was amplified from pSKBAD2 vector using the two oligonucleotides of SK001: 5′-TCCTTACATATGGCGTACAAACTTATCCTG and J165: 5′-ATGAAGCTTACAGAGCATTACGACCTTCCGTTACGGTGTAGGT, and cloned into pHYRSF1 using *Nde*I and *Hind*III sites, resulting in plasmid pKERSF3 [27,28]. This plasmid encodes GB1 under inducible T7 promoter with the C-terminal ligation tag of 'GRNAL'.

Plasmids for coexpression to produce a nicked MBP

The gene of maltose-binding protein (MBP)-N, the N-terminal fragment (residues 1–173) of *Escherichia coli* MBP with the C-terminal ligation tag sequence of 'CL' and E172K mutation was amplified from plasmid pFGRSF03 using the following two oligonucleotides of I393: 5′-AAACCATGGCTAAAATCGAAGAAGGTAAACTG and J141: 5′-ATGAAGCTTACAGACAGTTTTTATACTTGAACGCAT [25]. The amplified gene was cloned into pRSF-1b (Novagen) between *Nco*I and *Hind*III sites, resulting in pEMRSF13. MBP-C (residues 174–366) was cloned into pSABAD250 using *Nde*I and *Eco*RI sites after PCR amplification from pFGRSF03 as the template using the two primers of J144: 5′-AACATATGGGCCTGTACGACATTAAAGACG and J109: 5′-TTGGAATTCAGTCTGCGCGTCTTTCAGG [25,29]. The resulted plasmid of pEMBAD12 encodes MBP-C with 'GL' and K175L mutation at the N terminus and the C-terminal His-tag. MBP-C on pEMBAD12 can be induced by addition of L-arabinose.

Plasmid for the full-length MBP

The gene of *E. coli* MBP was amplified from plasmid pFGRSF03 [25] using the following two oligonucleotides of I393 and SZ031: 5′-GTCCAAGCTTAAGTCTGCGCGTC. The amplified gene was cloned into pRSF-1b (Novagen) between *Nco*I and *Hind*III sites, resulting in pBHRSF103.

Protein expression and purification

Expression, purification, and activation of *rOa*AEP1

The plasmid of pBHRSF184 bearing *Oa*AEP1 was transformed into chemically competent *E. coli* T7 SHuffle cells according to the manufacturer's instructions (*E. coli* K12 #C3026H; New England BioLabs, Inc., Ipswich, MA,

USA). The transformed cells were grown in Super Broth (SB) media supplemented with 10 $\mu g \cdot mL^{-1}$ kanamycin at 30 °C until OD_{600} reached 0.6. The cell culture was cooled down to 18 °C, followed by the induction of the fusion protein for 18 h with a final concentration of 0.1 mM isopropyl-β-D-thiogalactoside (IPTG). The induced cells were harvested by centrifugation for 10 min at 4000 ***g***, 4 °C. The harvested cell pellets were immediately resuspended in 20 mL of lysis buffer (50 mM Tris/HCl (pH 7.0), 150 mM NaCl, 0.1% Triton X100, and 1 mM EDTA) and disrupted by passing through Emulsiflex C-3 for 10 min at 15 000 psi, 4 °C. The cleared supernatant after centrifugation at 18 000 ***g*** for 60 min was diluted eight times with 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl. The cleared supernatant was subsequently loaded onto a 5 mL HisTrap HP column (GE Healthcare Life Sciences, Milwaukee, WI, USA) followed by washing with 200 mL of 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, and 30 mM imidazole. H_6-SMT3-*Oa*AEP1 fusion protein was eluted with a linear gradient of 30–300 mM imidazole. Fractions containing the fusion protein were collected and dialyzed against phosphate-buffered saline (PBS) buffer overnight at 8 °C. On the following day, the fusion protein was digested at 24 °C for 2 h by addition of a final concentration of 1 mM dithiothreitol (DTT) and 37 nM Ulp1 protease [25]. The reaction mixture was loaded onto HisTrap HP column to remove the N-terminally His-tagged SMT3 and Ulp1 protease. Flow-through fractions containing *rOa*AEP1 were collected and dialyzed overnight against 2 L of 20 mM HEPES/KOH (pH 8.0) and 150 mM NaCl. The purified *rOa*AEP1 was activated, following the previously reported protocol [22]. Briefly, after addition of final concentrations of 1 mM EDTA and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), pH of the protein solution was adjusted to 4.0 with 1 M sodium acetate for self-cleavage activation and incubated at 37 °C for 5 h. The cleaved active enzyme was stored at −74 °C for further usage.

Expression and purification of substrate proteins: GFP variants and GB1

Escherichia coli strain ER2566 was transformed using chemically competent cells with plasmids pITRSF4, pJTRSF50, pJTRSF55, pKERSF3, or pBHRSF208 and grown at 37 °C in LB medium supplemented with 25 $\mu g \cdot mL^{-1}$ kanamycin. When OD_{600} reached 0.6, the protein was induced with a final concentration of 1 mM IPTG and incubated for 3 h at 37 °C. The induced cells were harvested by 10-min centrifugation at 4000 ***g***, 4 °C. The cell pellet was resuspended in 20 mL of 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl and flash-frozen in liquid nitrogen for storage at −74 °C. The frozen cells were thawed and lysed at 15 000 psi for 10 min using Emulsiflex C-3. The cell debris were removed by centrifugation at 18 000 ***g***, 4 °C for 60 min. The supernatant was passed through 0.45-μm filter and loaded onto a 5 mL HisTrap HP column (GE Healthcare Life Sciences), which was pre-equilibrated with 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl. After washing with 50 mM sodium phosphate buffer (pH 8.0), 300 mM NaCl, and 30 mM imidazole, the protein of interest was eluted with a linear gradient of 30–300 mM imidazole in sodium phosphate buffer (pH 8.0) and 300 mM NaCl. The elution was dialyzed overnight against PBS buffer. The His-tagged SMT3 protein in the GFP fusion proteins was digested by Ulp1 protease in the presence of 1 mM DTT and removed by loading onto a 5 mL HisTrap HP column. The flow-through fractions containing the protein of interest (GFP with different C-terminal sequences) were dialyzed against 2 L of 20 mM Tris/HCl buffer (pH 8.0) and 150 mM NaCl overnight at 8 °C.

Expression, purification, and ligation of segmentally labeled MBP

Segmentally ^{15}N-labeled MBP was prepared by differential labeling using the time-delayed coexpression system, followed by ligation using *rOa*AEP1. First, differentially labeled two fragments of split MBP were expressed following the previously reported protocol using the dual vector system [27,28]. *Escherichia coli* strain ER2566 (New England Biolabs) was transformed with the two plasmids of pEMBAD12 and pEMRSF13 and plated on LB agar plates supplemented with 25 $\mu g \cdot mL^{-1}$ kanamycin and 100 $\mu g \cdot mL^{-1}$ ampicillin at 37 °C. About 50 mL of LB medium supplemented with a final concentration of 25 $\mu g \cdot mL^{-1}$ kanamycin and 100 $\mu g \cdot mL^{-1}$ ampicillin was inoculated with a single colony from the LB agar plate and incubated at 30 °C overnight with vigorous shaking at 200 r.p.m. The bacterial culture was diluted into 1.25 L of LB medium supplemented with the antibiotics and incubated at 37 °C. When OD_{600} reached 0.6, MBP-C was induced with a final concentration of 0.2% L-arabinose for 2 h at 37 °C. The cells were spun down for 20 min at 850 ***g*** in order to exchange the culture media. To remove the remaining inducer, the cell pellet was briefly washed by gently suspending with 15 mL of 100% ^{15}N-labeled M9-medium, followed by another centrifugation at 850 ***g***. The cell pellet was then resuspended in 1.8 L of 100% ^{15}N-labeled M9-medium supplemented with 25 $\mu g \cdot mL^{-1}$ kanamycin, 50 $\mu g \cdot mL^{-1}$ ampicillin, 50 $\mu g \cdot mL^{-1}$ carbenicillin, and a final concentration of 1 mM IPTG. MBP-N was induced for additional 3 h at 37 °C with shaking at 200 r.p.m. After the second induction, the cells were harvested with 10-min centrifugation at 5000 ***g*** and resuspended in 20 mL of 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl. The cell suspension was flash-frozen and stored for further purification. The frozen cells were thaw and lysed at 15 000 psi for 10 min using Emulsiflex C-3. The cell lysate was cleared by centrifugation at

18 000 ***g*** for 60 min at 4 °C. The supernatant was passed through 0.45-μm filter and loaded onto pre-equilibrated 5 mL HisTrap HP column (GE Healthcare Life Sciences) with 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl. After washing with 30 mM imidazole, 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl, the differentially labeled nicked MBP was eluted with a linear gradient of 30–300 mM imidazole in 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl. The elution fraction containing the nicked MBP was dialyzed overnight against 2 L of 20 mM Tris/HCl buffer (pH 8.0) and 150 mM NaCl. The protein solution was concentrated to 210 μL and flash-frozen in liquid nitrogen for storage at −74 °C.

Ligation of the nicked MBP was performed at 25 : 1 molar ratio for MBP and the enzyme. Final concentrations of 152 μM MBP and 6 μM activated *rOa*AEP1 were mixed in 0.5 mL volume for preparative ligation. The reaction mixture was transferred into a dialysis tube with MWCO 3.5 kDa and dialyzed overnight at 24 °C against 50 mM HEPES/KOH buffer (pH 7.0), 50 mM NaCl, 1 mM EDTA, and 0.5 mM TCEP. After overnight ligation reaction, the ligation mixture was further purified by anion exchange chromatography using MonoQ™ 5/50 GL (GE Healthcare Life Sciences). Only fractions containing the ligated MBP were dialyzed overnight at 4 °C against 1 L of 20 mM sodium phosphate buffer (pH 6.0). The ligated MBP was concentrated to 0.1 mM using an ultracentrifugation device. A final concentration of 0.3 mM maltose was added for NMR measurement.

Expression and purification of the uniformly ^{15}N-labeled MBP

The plasmid of pBHRSF103 bearing MBP was transformed into chemically competent *E. coli* strain ER2566 (New England Biolabs). One colony was inoculated in 5 mL of LB medium supplemented with 25 $\mu g \cdot mL^{-1}$ kanamycin and was grown at 37 °C for 7 h. The preculture was diluted into 50 mL of 100% ^{15}N-labeled M9-medium supplemented with 25 $\mu g \cdot mL^{-1}$ kanamycin and incubated overnight at 30 °C with 200 r.p.m. shaking. The overnight preculture was diluted with 450 mL of 100% ^{15}N-labeled M9 medium supplemented with 25 $\mu g \cdot mL^{-1}$ kanamycin. When OD_{600} reached 0.6, MBP was induced with a final concentration of 0.1 mM IPTG and incubated overnight at 24 °C. The induced cells were harvested by 10-min centrifugation, at 4000 ***g***, 4 °C. The cell pellet was resuspended in 10 mL of 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl and flash-frozen in liquid nitrogen for storage at −74 °C. The frozen cells were thawed and lysed at 15 000 psi for 10 min using Emulsiflex C-3. The cell debris were removed by centrifugation at 18 000 ***g***, 4 °C for 60 min. The supernatant was diluted with 30 mL of 20 mM Tris/HCl (pH 7.5), 200 mM NaCl, and 1 mM EDTA, and mixed with 15 mL of pre-equilibrated amylose resin beads (New England Biolabs). The supernatant and beads suspension were loaded onto an empty Econo-Pac chromatography column (BioRad, Hercules, CA, USA). The amylose beads were washed 5 times with 10 mL of 20 mM Tris/HCl (pH 7.5), 200 mM NaCl, and 1 mM EDTA. MBP was eluted 5 times with 5 mL of 20 mM Tris/HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA, and 10 mM maltose. All washing and elution fractions were analyzed on SDS/PAGE. Fractions containing a single band of MBP were collected and concentrated to a final concentration of 0.57 mM using an ultracentrifugation device. A final concentration of 1.71 mM maltose was added to the protein solution prior to the storage at −74 °C.

Cyclization assays with *rOa*AEP1

Five micromolar solution of each GFP variant with different C-terminal sequences of 'NGL', 'NAL', or 'NCL' in 50 mM HEPES/KOH buffer (pH 7.0) 50 mM NaCl, 1 mM EDTA, and 0.5 mM TCEP was mixed with a final concentration of 0.2 μM *rOa*AEP1 for backbone cyclization. The reaction mixture was incubated in an eppendorf tube at 24 °C for 2 h. The samples for SDS/PAGE analysis of backbone cyclization were taken after 30 min, 1 h, and 2 h of the reaction. Cyclized GFP migrates faster than linear GFP as previously reported and could be relinearized by thrombin digestion to confirm backbone cyclization [26].

Trans-ligation assays with *rOa*AEP1

Purified two substrate proteins were mixed at an equimolar ratio at a final concentration of 5 μM in the reaction buffer of 50 mM HEPES/KOH (pH 7.0), 50 mM NaCl, 1 mM EDTA, and 0.5 mM TCEP. Activated *rOa*AEP1 was then added at a final concentration of 0.2 μM at 24 °C in eppendorf tubes. The samples for SDS/PAGE analysis were taken after 1, 2 h, and 4 h of the reaction.

Quantification of the ligation yields

Time courses of the ligation/cyclization reactions were analyzed by SDS/PAGE. The samples taken at different time points were diluted with two-time SDS reducing loading buffer, and loaded on 12% or 18% SDS polyacrylamide gels. The SDS-gels were stained with Coomassie Blue R (GE Healthcare) and quantified using NIH IMAGEJ software (Bethesda, MD, USA), assuming that the dye binds equally to proteins [30].

NMR measurements

All the NMR spectra were recorded on Bruker Avance III HD spectrometer equipped with cryogenic probehead at ^{1}H frequency of 850 MHz. The NMR samples for segmentally and uniformly ^{15}N-labeled MBP were concentrated to

0.1 mM, 200 μL and 0.57 mM, 650 μL, respectively and transferred into Shigemi microcell tubes for NMR measurements.

Result and discussion

First, we recombinantly produced the codon optimized AEP1 from *O. affinis* (*rOa*AEP1) using *E. coli* overexpression system as a SUMO fusion protein replacing the reported ubiquitin fusion and purified as previously reported with slight modifications (see materials and methods) [22,25]. We discovered that the sequence of 'AL' or 'CL' at the P1′ and P2′ positions considerably improved the backbone cyclization efficiency of a model protein of GFP to > 95% with our recombinant *Oa*AEP1 (*rOa*AEP1; Fig. 2A,B) [26]. The cyclization efficiency of GFP with 'AL' was about 30% better than the previously reported 'GL' taken from kalata B1, which is different from the sequence of 'HV' of butelase 1 [21]. We decided to use either 'AL' or 'CL' because these sequences are more tolerant of amino acid changes at P2 position (data not shown). We next tested bimolecular ligation of two globular proteins, GFP and GB1, by *rOa*AEP1. For bimolecular ligation in *trans*, we added 'TRNAL' sequence at the C terminus of GB1 and 'GLP' at the N terminus of GFP, which is followed by a hexahistidine tag for purification (Fig. 2C). Ulp1 digestion was used to create the N-terminal glycine residue in the GFP for ligation. *rOa*AEP1 could ligate the two fragments at an equimolar ratio with about 50% yield although degradation of GB1 was evident presumably due to internal Asn residues (Fig. 2C). This is in agreement with the previous reports, suggesting that an excess of the reactants is necessary to achieve a higher yield [22–24]. To improve the ligation yield in *trans*, we decided to mimic backbone cyclization by producing a nicked protein where the N and C termini for ligation are in close proximity to each other as previously demonstrated with NCL [31]. Nicked proteins can be produced when two split fragments of a protein fold together inside *E. coli* cell after the coexpression (Fig. 3A) [28]. We used maltose-binding protein (MBP) as a model system. MBP was split within a loop into two fragments (residues 1–173 for MBP-N and residues 174–366 for MBP-C). We added a His-tag of 'EFH$_6$' at the C terminus of MBP-C after residue 366 for purification. MBP-N was cloned into pRSF vector under T7 promoter together with the C-terminal ligation tag of 'CL' after N173. We originally selected 'CL', because we thought that 'CL' could be better due to the thiol group. However, 'AL' was found to be as efficient as or slightly better than 'CL'

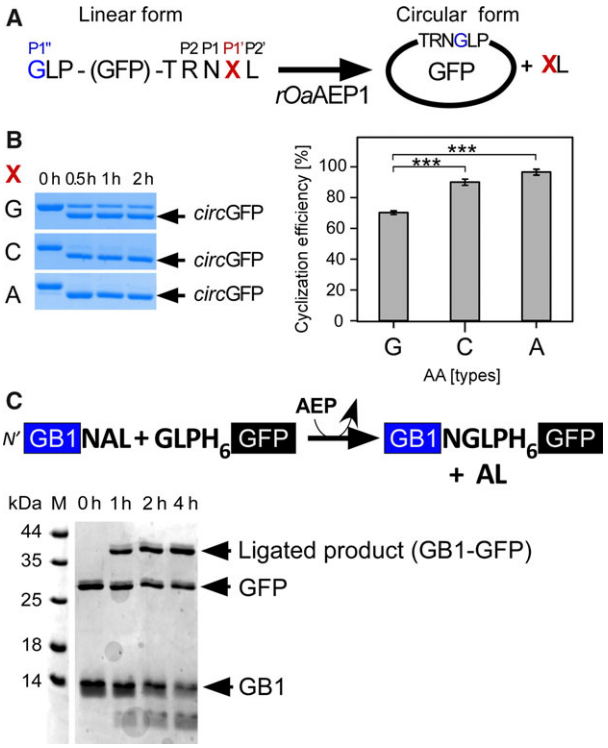

Fig. 2. Backbone cyclization and protein ligation in *trans* by *rOa*AEP1. (A) Schematic illustration of cyclization assay by *rOa*AEP1-catalyzed ligation. (B) SDS/PAGE analysis and quantification of the backbone cyclization of GFP substrates by *rOa*AEP1. 0, 0.5, 1, and 2 h stand for 0, 0.5, 1, and 2 h time points after addition of the enzyme. Arrows indicate the bands of circular GFP. *** indicates significance at the 0.001 probability level. (C) Schematic illustration and SDS/PAGE analysis of bimolecular protein ligation using GB1 and GFP by *rOa*AEP1.

(Fig. 2B). In addition, we replaced E172 with Lys. MBP-C with the C-terminal His-tag was cloned into pBAD vector under arabinose promoter for dual coexpression. In addition, K175 was replaced with Leu in MBP-C for the ligation by *rOa*AEP1 to achieve a higher yield (Fig. 3A). The transformed cells with these two plasmids could express MBP-C by induction with arabinose, followed by expression of MBP-N induced with an addition of IPTG (Fig. 3C). Because the two fragments could fold together inside *E. coli* cells and tightly associate, the nicked MBP was purified from soluble fraction owing to the folded conformation and the C-terminal His-tag (Fig. 3B) without any purification from insoluble fraction. Because the C terminus of MBP-N and the N terminus of MBP-C are in close proximity in the nicked protein, which is similar to the head-to-tail cyclization of cyclotides by AEPs, we expect that the ligation of these two fragments by *rOa*AEP1 is more favorable due to the

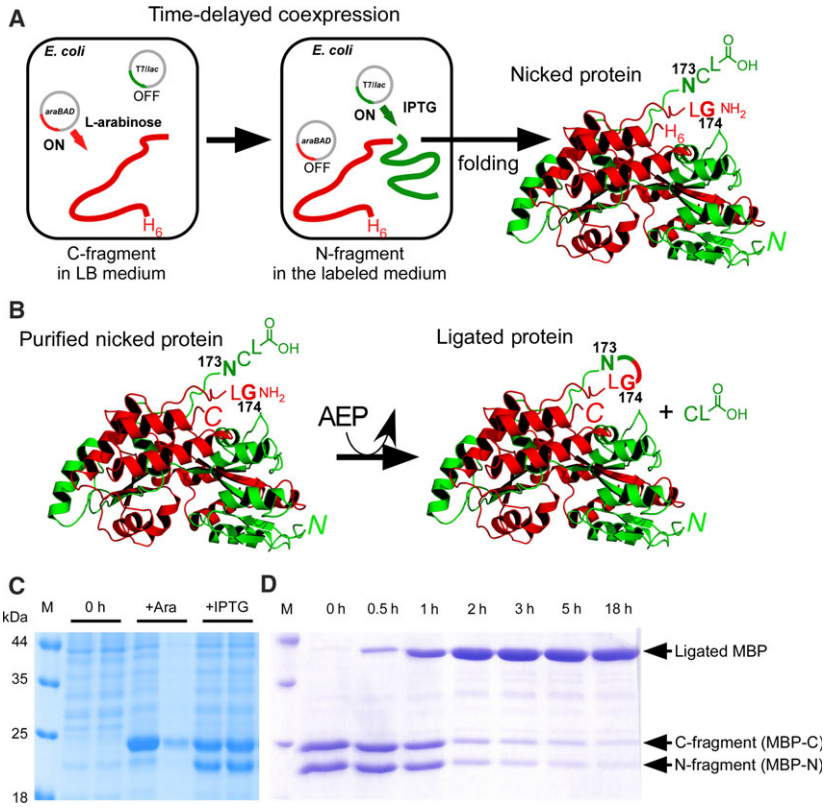

Fig. 3. Protein ligation of a nicked MBP by *rOa*AEP1. (A) Schematic drawing of production of a nicked MBP with the ligation tag for AEP using the dual coexpression system. (B) Schematic drawing of ligation of the nicked MBP by AEP to close the gap after purification. (C) SDS/PAGE analysis of overexpression of the N- and C-terminal fragments of MBP (MBP-N and MBP-C). '0 h' indicates before induction. '+Ara' indicates the first induction by arabinose for expressing MBP-C in the unlabeled medium. '+IPTG' indicates the second induction of MBP-N in the labeled medium after the first arabinose induction. (D) SDS/PAGE analysis of the time course of ligation of the nicked MBP by *rOa*AEP1. Reactions were performed at room temperature in the presence of 0.2 μM *rOa*AEP1 and 5 μM nicked MBP.

proximity effect than bimolecular *trans*-ligation without any affinity [31]. Moreover, we split MBP within an exposed loop so that the enzyme can access to the ligation site. More than 80% of ligation of the two fragments was indeed completed within 3 h without degradation (Fig. 3D). The ligated MBP has no additional amino acid insertion but only two residue changes of E172K and K175L near the ligation junction in the sequence because we exploited a natural 'NG' sequence in a loop of MBP.

One of the practical applications of protein ligation has been differential isotopic labeling of proteins to reduce NMR signal overlaps for NMR investigations of larger proteins and proteins with repeating sequences. EPL, PTS, and SrtA have been successfully used for segmental isotopic labeling of multidomain proteins containing self-contained domains [5,6,32–34]. However, segmental isotopic labeling of a single-domain globular protein is more challenging because individual split fragments of a globular domain alone are usually insoluble, requiring refolding steps of isolated fragments from insoluble fractions [9–11]. Hence, only PTS using split inteins have been used for a single-domain globular protein such as MBP, which requires labor-intensive optimization of the refolding conditions of the two precursors containing split intein fragments *in vitro* [9–11]. The shorter tag required for protein ligation by AEPs could facilitate segmental isotopic labeling of single-domain globular proteins because the smaller ligation tag of 'CL' or 'AL' is less likely to disturb the solubility of the split fragments to be ligated.

To demonstrate segmental isotopic labeling of MBP without any refolding step, we performed differential labeling of the two split MBP fragments (MBP-N and MBP-C) using the time-delayed coexpression as previously reported [28]. MBP-C was first induced in the unlabeled medium by arabinose, and then the medium was replaced with ^{15}N-labeled medium containing IPTG after a washing step. After the time-delayed dual-induction, the nicked MBP containing the associated ^{15}N-labeled MBP-N and unlabeled MBP-C was purified by immobilized metal ion affinity chromatography (IMAC). The gap between ^{15}N-labeled MBP-N and unlabeled MBP-C was closed with a peptide bond by protein ligation using *rOa*AEP1 (Fig. 3B,D). Figure 4 shows the comparison of HSQC spectra between uniformly ^{15}N-labeled MBP and the segmentally ^{15}N-labeled MBP by protein ligation using *rOa*AEP1, demonstrating effective segmental isotopic labeling of a single-domain globular protein without any refolding step. Both HSQC spectra recorded in the presence of

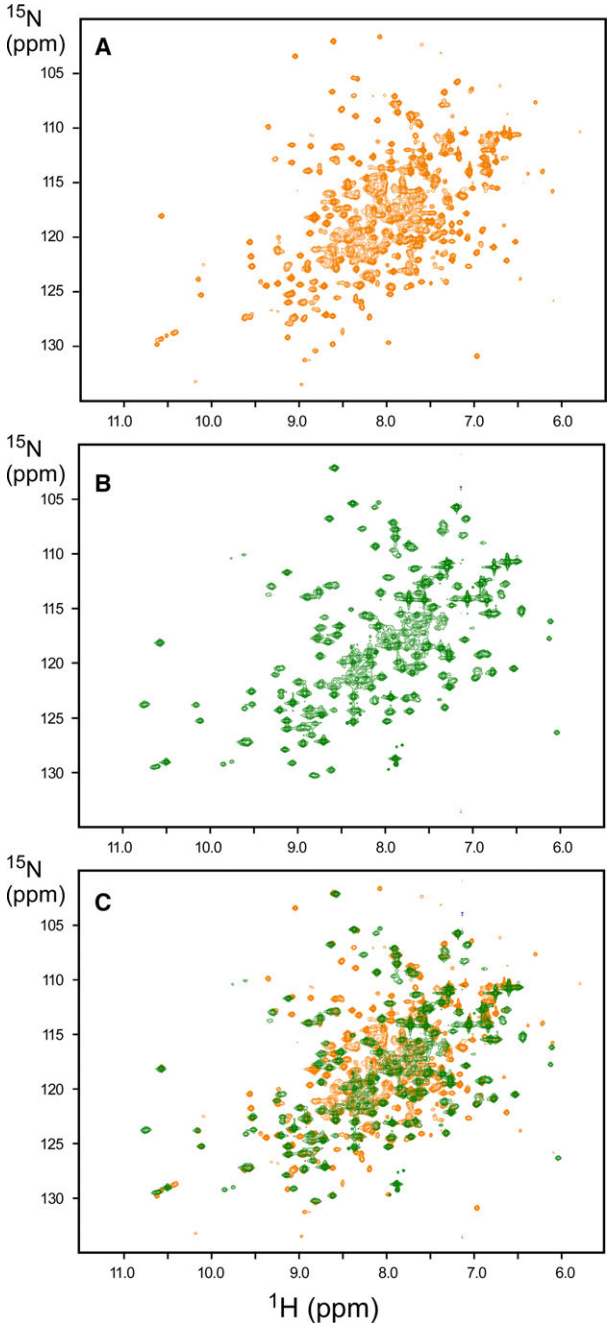

Fig. 4. (A) [^{1}H, ^{15}N]-HSQC spectrum of the uniformly ^{15}N-labeled full-length MBP. (B) [^{1}H, ^{15}N]-HSQC spectrum of segmentally [^{15}N, (1–173)]-labeled MBP produced by *rOa*AEP1. (C) A superposition of the two [^{1}H, ^{15}N]-HSQC spectra. The samples contain a threefold molar excess of maltose.

maltose can be overlaid without any shift or with slight shifts for visible peaks of MBP-N (residues 1–173), confirming that segmental labeled MBP produced by protein ligation is properly folded and capable of binding maltose (Fig. 4C). Some larger shifts are caused presumably by the sequence variations at the ligation site and the C-terminal His-tag used for purification of the ligated product. Importantly, the modification in the sequences required for protein ligation by *rOa*AEP1 is much smaller compared with PTS and SrtA approaches. The required sequences of 'NCL/NAL' and 'GL' for the C and N termini, respectively, are so small that AEP approach could be applicable to other nicked folded proteins when a split site is appropriately selected. Approximately, > 200 times better ligation efficiency of *rOa*AEP1 than SrtA and the shorter sequences required for ligation make recombinant *Oa*AEP1 produced from *E. coli* an attractive alternative [22]. In particular, *Oa*AEP1 has the different specificity from butelase 1. Recombinant production from *E. coli* could facilitate further protein engineering of *Oa*AEP1 for improving the enzymatic activity and creating desired specificities to broaden the application [35].

In summary, we demonstrated efficient protein ligation of a nicked protein with *rOa*AEP1 by making use of the proximity effect and the newly discovered recognition sequences of *Oa*AEP1 with better ligation efficiency. This approach could be widely applicable for segmental isotopic labeling of proteins particularly for single-domain globular proteins without labor-intensive refolding steps. When a nicked form of proteins can be produced by a dual-expression system, AEP-catalyzed ligation could be used for activating nicked proteins, e.g., inactive split toxic proteins that could not be expressed in *E. coli*. Less stringent proteolytic activity of AEPs due to the shorter recognition sequences might limit the application to relatively small globular domains. However, sequential combination of traceless ligation by PTS and enzymatic ligation by AEPs could be applied for production of more complex protein conjugates, incorporation of various biophysical probes in proteins, and production of full-length proteins in a post-translational manner.

Acknowledgements

This work is supported by the grants from the Academy of Finland (1137995, 1277335), TEKES (1311/31/2014), and Sigrid Jusélius Foundation. The NMR facility is supported by Biocenter Finland. The authors thank K. Elsner, B. Haas, S. Jääskeläinen, and E. Maschke for their technical help.

Author contributions

HI conceived and designed the project. KMM, JJT, IT and HI performed experiments and analyzed data. HI

and KMM wrote the paper with input from all authors.

References

1 Dawson PE, Muir TW, Clark-Lewis I and Kent SB (1994) Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779.

2 Evans CT, Benner J and Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**, 2256–2264.

3 Muir TW, Sondhi D and Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* **95**, 67705–67710.

4 Cao Y, Nguyen GKT, Tam JP and Liu CF (2015) Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation. *Chem Commun (Camb)* **51**, 17289–17292.

5 Minato Y, Ueda T, Machiyama A, Shimada I and Iwaï H (2012) Segmental isotopic labeling of a 140 kDa dimeric multi-domain protein CheA from *Escherichia coli* by expressed protein ligation and protein *trans*-splicing. *J Biomol NMR* **53**, 191–207.

6 Muona M, Aranko AS and Iwaï H (2008) Segmental isotopic labelling of a multidomain protein by protein ligation by protein *trans*-splicing. *ChemBioChem* **9**, 2958–2961.

7 Aranko AS, Oeemig JS and Iwaï H (2013) Structural basis for protein *trans*-splicing by a bacterial intein-like domain: protein ligation without nucleophilic side-chains. *FEBS J* **280**, 3256–3269.

8 Yamazaki T, Otomo T, Oda N, Kyogoku Y, Uegaki K, Ito N, Ishino Y and Nakamura H (1998) Segmental isotope labeling for protein NMR using peptide splicing. *J Am Chem Soc* **120**, 5591–5592.

9 Otomo T, Ito N, Kyogoku Y and Yamazaki T (1999) NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. *Biochemistry* **38**, 16040–16044.

10 Otomo T, Teruya K, Uegaki K, Yamazaki T and Kyogoku Y (1999) Improved segmental isotope labeling of proteins and application to a larger protein. *J Biomol NMR* **14**, 105–114.

11 Yagi H, Tsujimoto T, Yamazaki T, Yoshida M and Akutsu H (2004) Conformational change of H + -ATPase beta monomer revealed on segmental isotope labelling NMR spectroscopy. *J Am Chem Soc* **126**, 16632–166328.

12 Ciragan A, Aranko AS, Tascón I and Iwaï H (2016) Salt-inducible protein splicing in *cis* and *trans* by inteins from extremely halophilic archaea as a novel protein-engineering tool. *J Mol Biol* **428**, 4573–4588.

13 Yazawa K and Numata K (2014) Recent advances in chemoenzymatic peptide syntheses. *Molecules* **19**, 13755–13774.

14 Homandberg GA and Laskowski M (1979) Enzymatic resynthesis of the hydrolyzed peptide bond(s) in ribonuclease S. *Biochemistry* **18**, 586–592.

15 Chang TK, Jackson DY, Burnier JP and Wells JA (1994) Subtiligase: a tool for semisynthesis of proteins. *Proc Natl Acad Sci USA* **91**, 12544–12548.

16 Mao H, Hart SA, Schink A and Pollok BA (2004) Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **126**, 2670–2671.

17 Antos JM, Truttmann MC and Ploegh HL (2016) Recent advances in sortase-catalyzed ligation methodology. *Curr Opin Struct Biol* **38**, 111–118.

18 Dorr BM, Ham HO, An C, Chaikof EL and Liu DR (2014) Reprogramming the specificity of sortase enzymes. *Proc Natl Acad Sci USA* **111**, 13343–13348.

19 Lee J, McIntosh J, Hathaway BJ and Schmidt EW (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J Am Chem Soc* **131**, 2122–2124.

20 Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA and Craik DJ (2007) An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J Biol Chem* **282**, 29721–29728.

21 Nguyen GKT, Wang S, Qiu Y, Hemu X, Lian Y and Tam JP (2014) Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat Chem Biol* **10**, 732–738.

22 Harris KS, Durek T, Kaas Q, Poth AG, Gilding EK, Conlan BF, Saska I, Daly NL, van der Weerden NL, Craik DJ *et al.* (2015) Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat Commun* **6**, 10199.

23 Nguyen GKT, Qiu Y, Cao Y, Hemu X, Liu H-F and Tam JP (2016) Butelase-mediated cyclization and ligation of peptides and proteins. *Nat Protoc* **11**, 1977–1988.

24 Nguyen GKT, Kam A, Loo S, Jansson AE, Pan LX and Tam JP (2015) Butelase 1: a versatile ligase for peptide and protein macrocyclization. *J Am Chem Soc* **137**, 15398–15401.

25 Guerrero F, Ciragan A and Iwaï H (2015) Tandem SUMO fusion vectors for improving soluble protein expression and purification. *Protein Expr Purif* **116**, 42–49.

26 Iwaï H, Lingel A and Plückthun A (2001) Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-*Pfu*I intein from *Pyrococcus furiosus*. *J Biol Chem* **276**, 16548–16554.

27 Muona M, Aranko AS, Raulinaitis V and Iwaï H (2010) Segmental isotopic labeling of multi-domain and fusion proteins by protein *trans*-splicing *in vivo* and *in vitro*. *Nat Protoc* **5**, 574–587.

28 Züger S and Iwaï H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically

labeled proteins for NMR studies. *Nat Biotechnol* **23**, 736–740.
29 Aranko AS, Oeemig JS, Kajander T and Iwaï H (2013) Intermolecular domain swapping induces intein-mediated protein alternative splicing. *Nat Chem Biol* **10**, 616–622.
30 Schneider CA, Rasband WS and Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **7**, 671–675.
31 Beligere GS and Dawson PE (1999) Conformationally assisted protein ligation using C-terminal thioester peptides. *J Am Chem Soc* **121**, 6332–6333.
32 Freiburger L, Sonntag M, Hennig J, Li J, Zou P and Sattler M (2015) Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *J Biomol NMR* **63**, 1–8.
33 Skrisovska L and Allain FH-T (2008) Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRMs of Npl3p and hnRNP L. *J Mol Biol* **375**, 151–164.
34 Liu D, Xu R and Cowburn D (2009) Segmental isotopic labeling of proteins for nuclear magnetic resonance. *Meth Enzymol* **462**, 151–175.
35 Yang R, Wong YH, Nguyen GKT, Tam JP, Lescar J and Wu B (2017) Engineering a catalytically efficient recombinant protein ligase. *J Am Chem Soc*. doi:10.1021/jacs.6b12637.