Salt-inducible Protein Splicing in *cis* and *trans* by Inteins from Extremely Halophilic Archaea as a Novel Protein-Engineering Tool

Annika Ciragan, A. Sesilja Aranko, Igor Tascon and Hideo Iwaï

Research Program in Structural Biology and Biophysics, *Institute of Biotechnology, University of Helsinki, P.O. Box 65, Helsinki, FI-00014, Finland*

Correspondence to Hideo Iwaï: *hideo.iwai@helsinki.fi*
http://dx.doi.org/10.1016/j.jmb.2016.10.006
Edited by S. Koide

Abstract

Intervening protein sequences (inteins) from extremely halophilic haloarchaea can be inactive under low salinity but could be activated by increasing the salt content to a specific concentration for each intein. The halo-obligatory inteins confer high solubility under both low and high salinity conditions. We showed the broad utility of salt-dependent protein splicing in *cis* and *trans* by demonstrating backbone cyclization, self-cleavage for purification, and scarless protein ligation for segmental isotopic labeling. Artificially split MCM2 intein derived from *Halorhabdus utahensis* remained highly soluble and was capable of protein *trans*-splicing with excellent ligation kinetics by reassembly under high salinity conditions. Importantly, the MCM2 intein has the active site residue of Ser at the +1 position, which remains in the ligated product, instead of Cys as found in many other efficient split inteins. Since Ser is more abundant than Cys in proteins, the novel split intein could widen the applications of segmental labeling in protein NMR spectroscopy and traceless protein ligation by exploiting a Ser residue in the native sequences as the +1 position of the MCM2 intein. The split halo-obligatory intein was successfully used to demonstrate the utility in NMR investigation of intact proteins by producing segmentally isotope-labeled intact TonB protein from *Helicobacter pylori.*

Introduction

Halophiles thrive under high salinity conditions by deploying various approaches to counter the increased osmotic pressure caused by high salinity [1]. Halo-adaptation of halophilic proteins at the molecular level is a fascinating research topic [2,3]. Inteins catalyzing protein splicing are prevalent in halophilic organisms [4,5]. Protein splicing catalyzed by inteins is a post-translational auto-processing modification in which an intervening sequence (intein) excises itself from the host precursor protein, simultaneously ligating the two flanking sequences (exteins) with a peptide bond (Fig. 1a) [6,7]. This process is self-catalytic, requiring neither additional co-factors nor energy, and has become an important protein-engineering tool. A number of inteins have been identified in extremely halophilic archaea, and some of them have been reported to be inactive in *Escherichia coli* [5,8]. There has been growing interest in controlling protein splicing using small molecules, temperature, lights, redox potential, and domain swapping to activate the host protein functions for developing new biotechnological applications (Fig. 1b) [9–13]. Conditional protein splicing (CPS) allows us, for example, to have spatial and temporal control of the host proteins of inteins [14]. Protein-fragment complementation by split inteins has been a powerful tool for both *in vivo* protein engineering and *in vitro* protein ligation, because split inteins can ligate polypeptide chains via protein *trans*-splicing (PTS) (Fig. 1c) [15–17]. Therefore, there have been many attempts to identify novel split inteins with different lengths and robust splicing activity for various *in vivo* and *in vitro* applications [18–23]. However, artificially splitting inteins to create robust split inteins has not been very successful, because splitting a folded protein into two pieces makes them less soluble due to the increase in exposed hydrophobic area caused by disrupting the three-dimensional structure [24,25]. Introducing a split site too close to the termini could

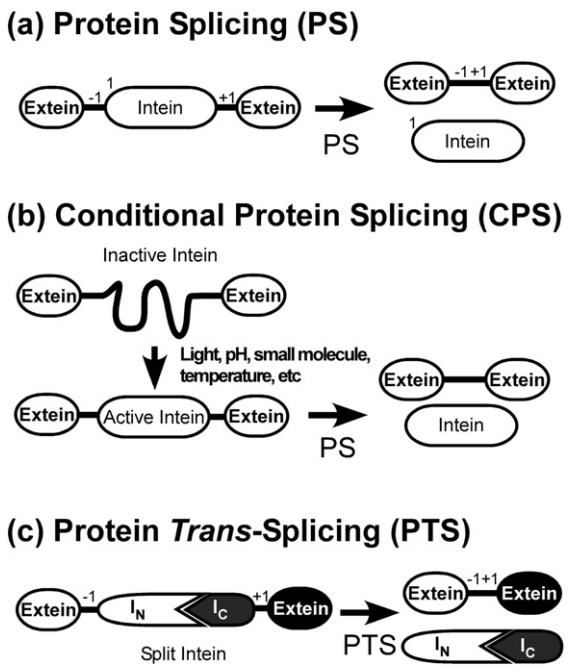

Fig. 1. Protein splicing in *cis* and *trans*. (a) Protein splicing ligates the flanking sequences between the −1 and +1 positions at the splicing junctions, concomitantly excising out the intein after protein translation and folding. (b) CPS in *cis*. Intein is unfolded or inactive before the induction of protein splicing by, for example, light, pH, small molecule ligand, salts, redox potential, or temperature. (c) Protein ligation by protein *trans*-splicing (PTS) uses split intein fragments (Int_N and Int_C) fused with the target fragments (exteins).

also induce partial splicing activity, resulting in premature cleavages [19,26]. The most widely used split inteins for biotechnological applications have been naturally occurring split inteins found in cyanobacteria [18,23,27].

Here, we report salt-inducible protein splicing in *cis* and *trans* using inteins from extreme halophilic archaea that are inactive under low salinity but can be activated under high salinity as a novel CPS system for backbone cyclization, protein purification, and scarless protein ligation for segmental isotopic labeling of an intact protein. We designed a novel, highly efficient split MCM2 intein from *Halorhabdus utahensis* for PTS and demonstrated its practical utility for segmental isotopic labeling of intact TonB protein from *Helicobacter pylori* by highlighting the importance of segmental isotopic labeling with the native sequence for NMR investigation.

Results

To expand the practical applications of protein ligation using PTS, it has been critical to identify robust split inteins with better ligation efficiency and wider substrate specificity (various junction sequences), such as the naturally occurring split DnaE intein found in *Nostoc punctiforme* (*Npu*DnaE intein) [18,20–22]. Inteins are unique enzymes having only a single turnover because substrates (exteins) are covalently bound to the enzyme (intein). We previously reported the comparison of protein-splicing activity among relatively small inteins from various organisms to identify promiscuous inteins bearing different junction sequences with high splicing efficiencies [28]. We noticed that some inteins did not spontaneously splice when they were overexpressed in *E. coli* using a model system of the B1 domain of IgG binding protein G (GB1) as the flanking exteins (Fig. 2a) [28]. Among them, two inteins from *Halobacterium salinarum* NRC-1 (ATCC 700922), *Hsa*PolII and *Hsa*CDC21 inteins, caught our attention. *H. salinarum* NRC-1 has an optimal growth at 4.3 M NaCl and an intercellular potassium concentration of 4.6 M and produces intein-less mature host proteins where inteins are inserted [8,28]. We therefore tested protein-splicing activities in the presence of 4 M NaCl (Fig. 2b). Both *Hsa* inteins could produce the expected spliced product *in vitro* only at a high salt concentration, suggesting that the solution condition is responsible for protein splicing rather than the host protein contexts, namely the substrate (junction sequences) specificity of the inteins (Fig. 2b and c). *Cis*-splicing of these inteins under high salinity condition has also been recently reported [8,29]. The required high salt concentration indicates that these inteins from extreme halophilic archaea are strictly halophilic (or halo-obligatory) instead of halo-tolerant (Fig. 2b and c). This observation prompted us to screen other inteins from different halophilic organisms in order to identify inteins that would be useful for protein ligation, with robust splicing activity and less side-products by cleavages, because both *Hsa* inteins produced higher amounts of cleaved products than the spliced product (Fig. 2b and c). We discovered that the MCM2 intein from *H. utahensis* (*Hut*MCM2 intein) possesses better protein expression and splicing activity in the presence of high salt concentrations than *Hsa* inteins (Fig. 2b). Notably, the *Hut*MCM2 intein could be activated at a lower concentration of NaCl without it producing notable quantities of side-products using a model system (Fig. 2c and Supplementary Figs. 1 and 2**)**.

Protein purification by salt-induced cleavage

Canonical protein-splicing reaction involves the four concerted reaction steps of (1) N-S(O) acyl shift, (2) transesterification, (3) Asn cyclization, and (4) S(O)-N acyl shift [30]. A mutation of Asn to Ala at the last residue of the inteins has been introduced to create self-cleavable affinity tags for efficient protein purification by halting the protein-splicing reaction after the first step of N-S acyl shift, thereby making

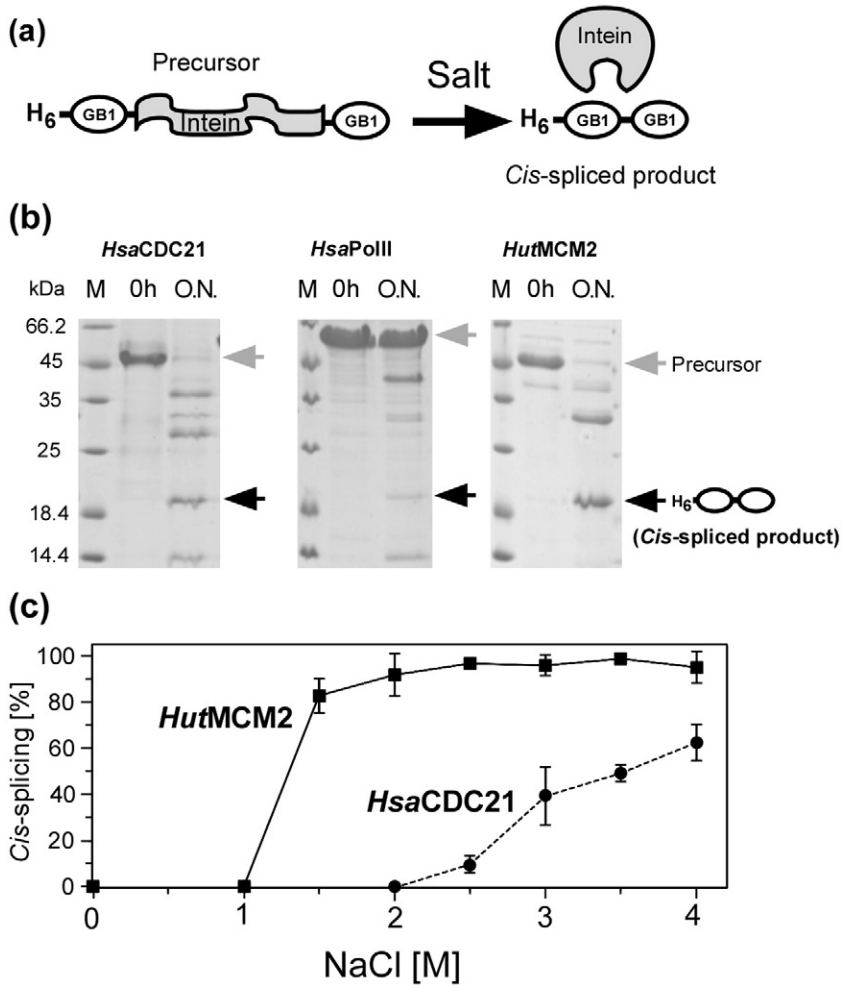

Fig. 2. Salt-dependent CPS. (a) Salt induction of protein *cis*-splicing using GB1 as exteins. (b) Protein *cis*-splicing of three inteins from extreme halophilic archaea. The samples were taken immediately after induction with 4 M NaCl (0 h) and after overnight incubation (O.N.) and were analyzed by SDS-PAGE. Grey and black arrows indicate the precursor and spliced product, respectively. (c) *Cis*-splicing efficiency *versus* NaCl concentrations for the *Hut*MCM2 intein (solid line) and the *Hsa*CDC21 intein (dotted line). The mean values (filled circles and squares) and errors were obtained from triplicated measurements.

the intermediate thioester susceptible to cleavage by reducing agents such as DTT or by hydrolysis (Fig. 3a) [31]. This self-cleavable affinity tag system has been commercially available as IMPACT™ system (New England Biolabs) [31]. However, premature cleavages without any cleaving agents due to hydrolysis of the thioester intermediate have been observed, depending on the chemical stability of thioester intermediate [32,33]. This suggests that the three-dimensional structure of the intein is sufficient to promote the first step of N-S acyl shift. The halo-obligatory *Hut*MCM2 intein is inactive under low salinity conditions due to unfolded conformation as observed by NMR spectroscopy (data not shown). Therefore, it is implausible that the unstructured halophilic intein could induce the first step of N-S migration under low salinity conditions, thereby preventing premature cleavages frequently observed for other inteins. To verify this hypothesis, we introduced two Ala mutations at the last residue and at the +1 position of *Hut*MCM2 intein and a C-terminal purification tag (octa-histidine tag) in the fusion protein (Fig. 3a). The fusion protein could be easily purified with the C-terminal His-tag without any premature cleavages. The self-cleavage of the fusion protein was then induced by the addition of a high concentration of NaCl (Fig. 3b). The intein tag bearing the C-terminal His-tag could be subsequently removed by an additional purification step using immobilized metal chelating (IMAC) chromatography (Fig. 3b). This result confirms that the variant of a halo-obligatory intein bearing the mutations at the C-terminal junction

can be used as a novel intein-mediated protein purification system exploiting the salt-induced cleavage. Importantly, the salt-controlled structure formation of halo-obligatory inteins can circumvent any premature cleavages without the optimization of protein expression conditions, for example, the expression temperature [31–33]. Despite the lower cleavage efficiency, increasing salt concentration without any reducing agents was sufficient to cleave the fusion protein (Fig. 3c). This cleavage system might be beneficial to proteins containing disulfide bonds.

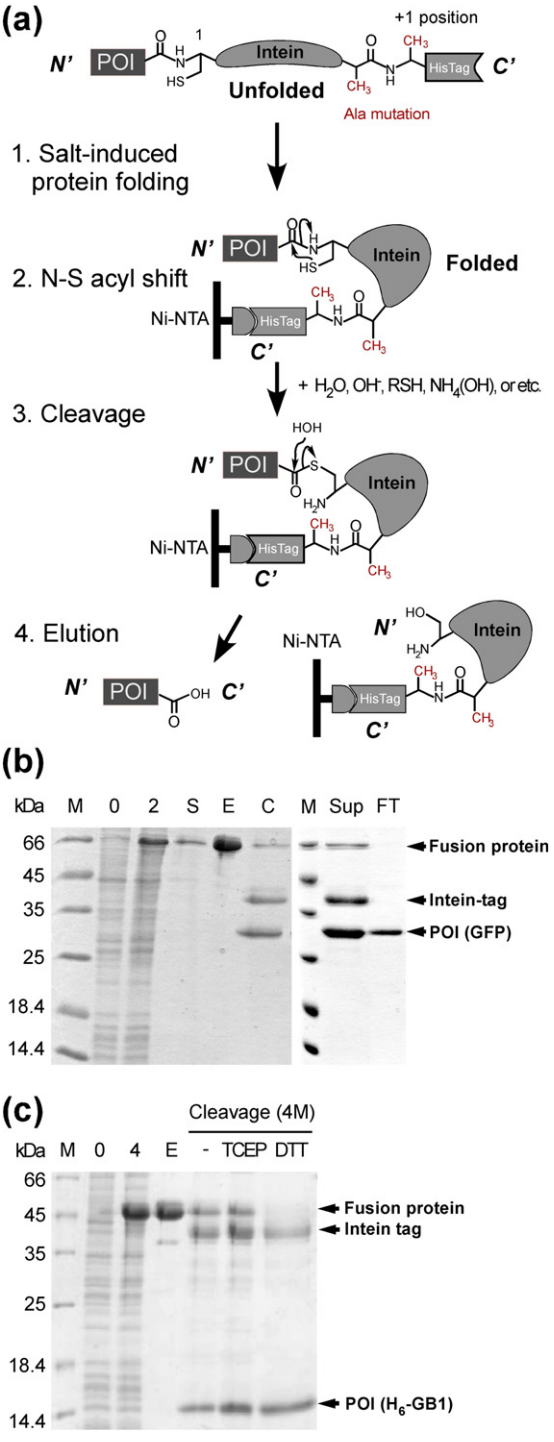

Fig. 3. Protein purification by salt-induced self-cleavage. (a) Cleavage reaction by salt-inducible self-cleavage using *Hut*MCM2 intein with the double mutations of Ala at the last residue and at the +1 position of the intein; Step 1, Salt-induced protein folding; Step 2, N-S acyl shift by the folded intein fragment; Step 3, Hydrolysis by hydroxyl or thiol reagents; Step 4, Elution of the cleaved product. (b) SDS-PAGE analysis of the expression and purification steps. Lane M, molecular marker; lane 0, total cell lysate before protein induction; lane 2, 2 h after induction; lane S, supernatant after centrifugation of cell lysate; lane E, elution from Ni-NTA column; lane C, after induced cleavage by NaCl incubation. Right panel: SDS-PAGE analysis of second purification step using the Ni-NTA column to remove the intein tag and unreacted fusion protein. Lane M, molecular marker; lane Sup, reaction mixture before loading; lane FT, flow-through fraction from the Ni-NTA column. (c) SDS-PAGE analysis of the protein expression, purification, and cleavage. Lane M, molecular marker; lane 0, total cell lysate before protein induction; lane 4, 4 h after induction; lane E, elution from the Ni-NTA column; lane -, cleavage in 4 M NaCl without any reducing agent; lane TCEP, cleavage in 4 M NaCl with 0.5 mM TCEP; lane DTT, cleavage in 4 M NaCl with 1 mM DTT.

Salt-induced protein cyclization

The next question was whether we could develop a split intein system from a halo-obligatory intein for salt-inducible protein ligation by PTS. First, we tested *in vitro* protein cyclization by using a split halo-obligatory intein because it is a simpler intramolecular splicing reaction. As reported previously, we created a fusion protein by introducing a circular permutation of an intein-containing precursor bearing green fluorescent protein (GFP) as the target protein (Fig. 4) [34]. Backbone cyclization has been shown to increase the stabilities of peptides and proteins without any changes in the primary structure when both the N- and C-termini are structurally in close proximity [35,36,38]. However, intein-mediated protein cyclization using a split intein spontaneously produces a circular protein already within the *E. coli* cells [34,36]. Therefore, a purification tag needs to be incorporated in the target protein/peptide for the convenient purification. Such a purification tag might not be always desirable or possible to incorporate e.g., for smaller proteins and peptides. Inducible protein/peptide cyclization would be able to facilitate the purification step of circular proteins/peptides by implementing a purification tag only in the precursor but not in the target (Fig. 4). Backbone cyclization can be induced *in vitro* by salt induction after the purification step of the precursor fusion protein (Fig. 4a). For salt-inducible backbone cyclization,

we tested *Hsa*CDC21 intein that was split after residue 120 and fused with GFP because the intramolecular reaction can be less prone to side reactions (Fig. 4a). The split site was selected because the site corresponds to the naturally split site of the *Npu*DnaE intein as predicted from the sequence alignment [25]. The full-length precursor protein was purified without any *in vivo* cyclization reaction from *E. coli* using a His-tag at the N-terminus of the precursor protein. Backbone cyclization using the precursor protein was then induced *in vitro* by addition of a final concentration of 3.5 M NaCl. Figure 4b shows the salt-induced *in vitro* backbone cyclization of GFP bearing a thrombin cleavage site. Interestingly, only circular GFP was produced by the salt induction, judging from the SDS-PAGE analysis. The circular form was relinearized by the digestion of the GFP using thrombin to confirm the cyclization, which migrated slower than the circular form as previously reported (Fig. 4b and c) [34]. This result demonstrated the feasibility of backbone cyclization of proteins/peptides by salt induction using halo-obligatory split inteins.

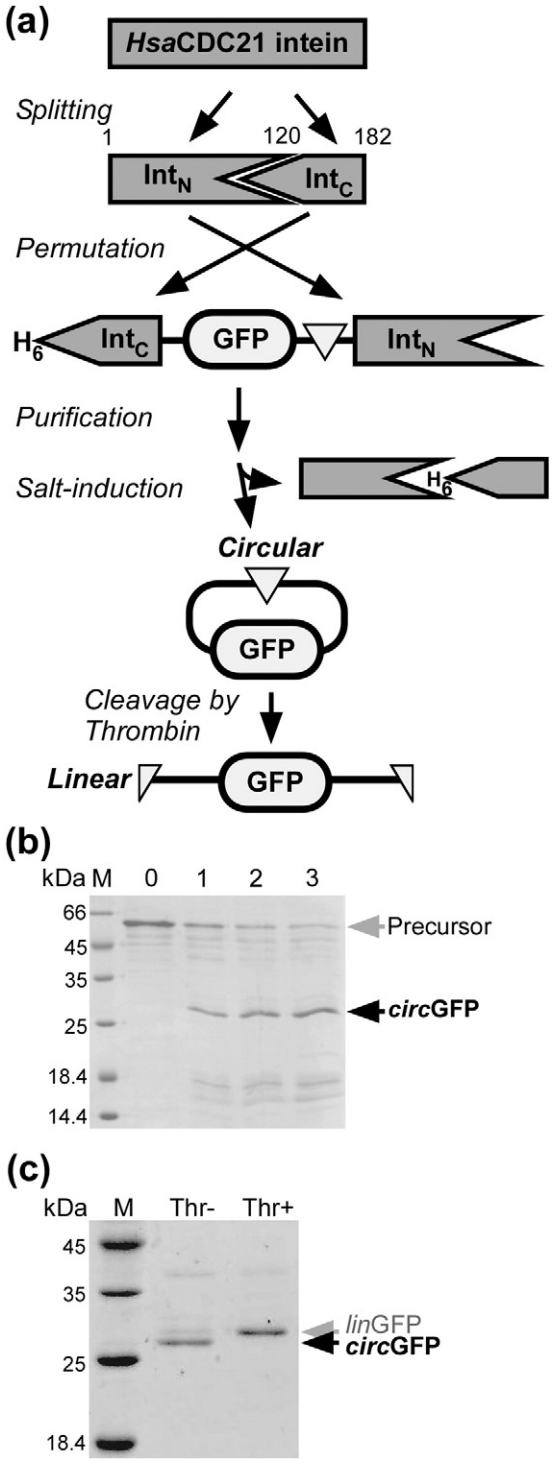

Fig. 4. Protein cyclization by salt-inducible protein splicing using *Hsa*CDC21 intein. (a) Schematic drawing of protein cyclization by inducible protein splicing. The inverse triangle indicates a specific cleavage site of thrombin used for the backbone linearization. (b) Time course of cyclization of GFP in the presence of 3.5 M NaCl. Lanes 0, 1, 2, and 3 indicate 0 h, 1 h, 3 h, and 19 h after the salt-induction, respectively. Grey and black arrows indicate the precursor and cyclized product, respectively. (c) Linearization of circular GFP by thrombin digestion. Thr- and Thr+ indicate without or with thrombin digestion, respectively. Black and grey arrows indicate circular and linear GFP, respectively.

Salt-induced protein ligation by PTS

Encouraged by the result from backbone cyclization, we next tested bimolecular protein ligation by salt-inducible PTS using the *Hut*MCM2 intein. This intein was split into two halves after residue 117, similar to *Hsa*CDC21 intein [25]. Each of the split halves was fused with a GB1 domain as an extein for testing PTS (Fig. 5a). *Trans*-splicing kinetics of the artificially split *Hut*MCM2 intein was $4.5 \pm 0.6 \times 10^{-4}$ (s^{-1}) for the model system of GB1s, with one N-terminal and three C-terminal native junction sequences ("R" and "SED", respectively) including Ser at the +1 position (Fig. 5a). This ligation kinetics of the split *Hut*MCM2 intein is as good as that of the well-characterized robust split *Npu*DnaE intein with the same model system of GB1, even though *Npu*DnaE has more nucleophilic Cys than Ser at the +1 position [18,19]. Because the +1 position of inteins will remain in the ligated product, the +1 residue often results as a scar in the ligated protein by PTS. Since Ser is more abundant than Cys, *Hut*MCM2 intein having Ser at the +1 position can be widely used for various applications by exploiting Ser residues in intact proteins. Moreover, the split *Hut*MCM2 intein was more soluble than the widely used *Npu*DnaE and

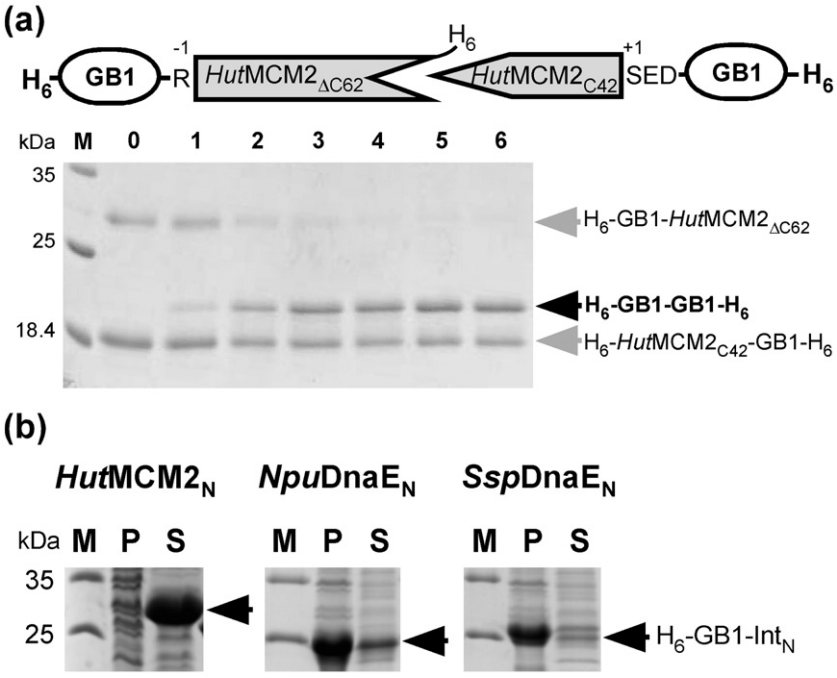

Fig. 5. PTS using the halo-obligatory split *Hut*MCM2 intein. (a) *In vitro* protein ligation by inducible PTS using a model system of GB1. Lane 0, before salt-induction; lanes 1, 2, 3, 4, 5, and 6 indicate samples taken 10 min, 30 min, 1 h, 2 h, 3 h, and 6 h after the 4 M NaCl induction, respectively. (b) Comparison of the soluble and insoluble fractions of the N-terminal precursor (Int_N fused with the identical N-extein of GB1). Int_N was from *Hut*MCM2 (*Hut*MCM2$_{\Delta C62}$), *Npu*DnaE, or *Ssp*DnaE inteins. M, P, and S stand for marker, pellet, and supernatant, respectively.

*Ssp*DnaE inteins when they were fused with an identical extein (Fig. 5b and Supplementary Fig. 3a). This is presumably because of the high solubility of strictly halophilic inteins as observed for other halophilic proteins [37,39]. Salt-inducible split inteins derived from halo-obligatory inteins could thus overcome the solubility issues of many artificially split inteins, which have been used for applications including segmental isotopic labeling [40].

Scarless segmental isotopic labeling of TonB from *H. pylori*

To demonstrate the practicality of salt-inducible inteins, we used salt-induced *in vitro* protein ligation for segmental isotopic labeling. We chose the TonB protein from *H. pylori* (*Hp*TonB) for protein ligation, which was conducted using a split *Hut*MCM2 intein system. TonB protein is a membrane-anchored protein essential for TonB-dependent outer membrane transporters, consisting of a transmembrane (TM) region, a Pro-rich region, and a C-terminal domain (CTD) (Fig. 6a) [41–43]. The dissected CTDs of TonB proteins have a well-folded globular domain. However, significant structural differences in the three-dimensional structures of *E. coli* CTD have been reported for differently dissected CTDs with various lengths [41–43]. Therefore, the three-dimensional structure of the CTD of the TonB protein in the full-length context could shed light on how intact TonB protein functions. The Pro-rich region is highly flexible as evidenced by severe NMR signal overlaps, which could hinder the structural characterization of the full-length *Hp*TonB in detail by both NMR spectroscopy and crystallography (see below). *Hp*TonB was split between the Pro-rich region and the CTD (after residue 154) in order to use Ser155 residue as the +1 position of *Hut*MCM2 intein and was fused with two halves of a split *Hut*MCM2 intein to create N- and C-terminal precursors (Fig. 6b and c). For the N-terminal precursor, the N-terminal fragment containing the Pro-rich region (residues 36–154) of *Hp*TonB was fused with the N-terminal fragment (residues 1–117) of *Hut*MCM2 intein (*Hut*MCM$_{\Delta C62}$) together with the C-terminal octa-histidine tag (Fig. 6b and c). For the C-terminal precursor, CTD (residues

Fig. 6. (a) Primary structure and domains of *Hp*TonB. TM region is highlighted in grey. Pro-rich region and CTD are indicated by closed rectangles and broken rectangles in grey, respectively. (b) Sequence alignment of the junction regions of *Hp*TonB, two precursors, and the expected ligation product. The split site used for segmental labeling is indicated by "/". (c) Procedure for protein ligation of *Hp*TonB using inducible PTS for segmental isotopic labeling. Residue numbers of *Hp*TonB are shown on the top. (d) Protein ligation of *Hp*TonB using split *Hut*MCM2 intein with Arg at the −1 position. Lane N, purified N-terminal precursor before mixing; lane C, purified C-terminal precursor before mixing; lane 0 h, immediately after mixing the N- and C-terminal precursors in 3.5 M NaCl; lane 16 h -, after 16-h dialysis in 0 M NaCl; lane 16 h +, after 16-h dialysis in 3.5 M NaCl; lane 22 h -, after 22-h dialysis in 0 M NaCl; lane 22 h +, after 22-h dialysis in 3.5 M NaCl; lane 39 h -, after 39-h dialysis in 0 M NaCl; lane 39 h +, after 39-h in 3.5 M NaCl; lane LP, the ligation product after the IMAC purification.

155–285) of *Hp*TonB was fused with the C-terminal fragment (residues 145–186) of *Hut*MCM2 intein (*Hut*MCM2$_{C42}$) together with the N-terminal hexahistidine tag for ^{15}N-labeling (Fig. 6b and c). Both N- and C-terminal precursors were cloned in the expression vectors under the control of a T7 promoter and were

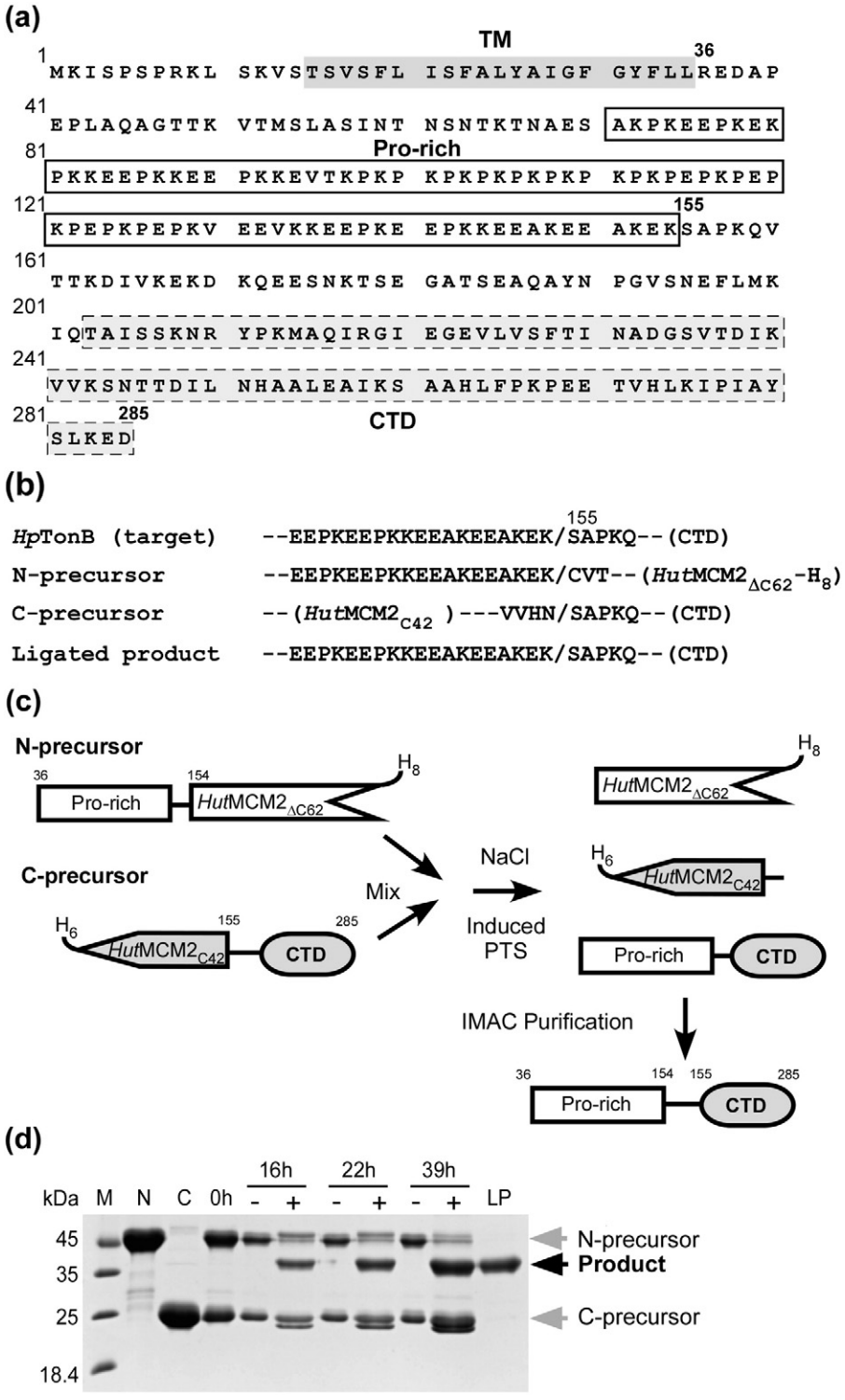

Fig. 6 (*legend on previous page*)

expressed in *E. coli*. The precursors were individually purified by IMAC without any solubilization steps (Fig. 6c and Supplementary Fig. 3b). The two purified precursors in phosphate-buffered saline (PBS) were mixed at a final concentration of ~0.1 mM in a total volume of 4.5 mL. The solution mixture was dialyzed against 1 L of 3.5 M NaCl, 0.5 M NaPi, and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) at pH 7.0 for 1 to 2 days at room temperature, followed by a buffer exchange to PBS buffer before further purification. The ligated product was further purified by another round of IMAC because all of the unreacted precursors and excised split intein fragments remained tagged with poly-histidine tags and could be efficiently removed. The flow-through fractions from IMAC were collected and analyzed by SDS-PAGE (Fig. 6d). Typically, the ligated *Hp*TonB(36–285) with >95% purity (~3 mg of the ligated product from 0.5 L of each cell culture) was obtained after 1 to 2 days of dialysis. The segmentally labeled sample was also confirmed by mass spectrometry (Supplementary Fig. 4). The ligation efficiency was estimated to be ~30% after a 1-day dialysis for *Hp*TonB with the native sequence (Lys at residue 154). The ligation efficiency was better (>~80% after a 2-day dialysis) when the native −1 residue (Arg) of *Hut*MCM2 intein was used instead of Lys at the −1 position (residue 154 in *Hp*TonB). The lower efficiency and slower kinetics compared to the model system suggest that both the junction sequences (the −1 and +2 positions) and the target protein still significantly influence the final yield, as has been previously observed with other inteins [18,22]. However, the precursors and split intein fragments were efficiently removed by purification tags in the split inteins, despite the lower efficiency for the ligation of intact *Hp*TonB (Fig. 6c and d). Importantly, the native sequence of *Hp*TonB was produced without any scar because the "KS" sequence in *Hp*TonB was utilized as the −1 and +1 positions of *Hut*MCM2 intein for protein ligation, thereby resulting in the native sequence of *Hp*TonB at the splicing junction.

Figure 7 shows the comparison of the uniformly ^{15}N-labeled *Hp*TonB(36–285) and the segmentally [^{15}N, (155–285)]-labeled intact *Hp*TonB(36–285) produced by salt-induced PTS. The uniformly labeled sample shows the heavily overlapped strong NMR signals within a narrow range around 8–8.5 ppm in the [^{1}H, ^{15}N]-heteronuclear single quantum coherence (HSQC) spectrum due to the highly flexible Pro-rich region (Fig. 7a). On the other hand, the [^{1}H, ^{15}N]-HSQC spectrum of the segmentally isotope-labeled sample presents the well-dispersed peaks with homogenous intensities originating only from the CTD (residues 155–285) of *Hp*TonB (Fig. 7b). Moreover, the ^{1}H 1D spectrum of the segmentally labeled sample still exhibits strong ^{1}H signals from the Pro-rich region, validating that the flexible Pro-rich region is not labeled with ^{15}N due to the segmental isotopic labeling (Fig. 7c). The segmentally isotope-labeled *Hp*TonB with the native sequence could thus facilitate the detailed NMR analysis of CTD in the presence of the flexible Pro-rich region.

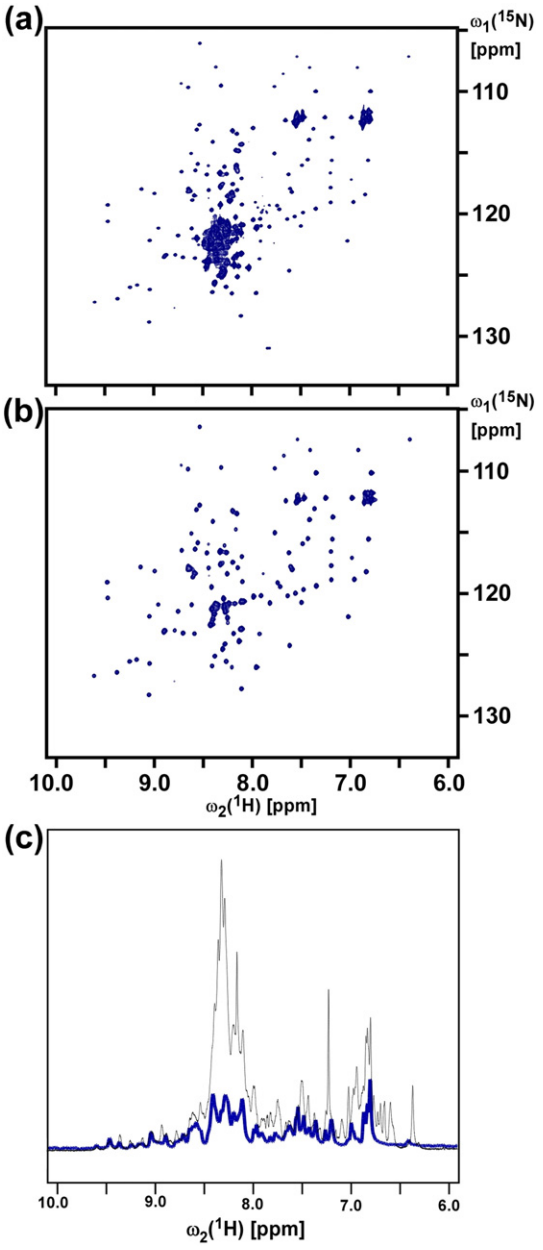

Fig. 7. [^{1}H, ^{15}N]-HSQC spectra from (a) uniformly ^{15}N-labeled *Hp*TonB(36–285) and (b) segmentally [^{15}N, (155–285)]-labeled *Hp*TonB(36–285). (c) Comparison of 1D ^{1}H–NMR spectrum (thin line) and 1D ^{1}H projection from [^{1}H, ^{15}N]-HSQC spectrum of the identical sample of segmentally [^{15}N, (155–285)]-labeled *Hp*TonB(36–285).

The comparison between CTD and the full-length *Hp*TonB without TM

To evaluate the structural influences of the Pro-rich region on the CTD of *Hp*TonB, we determined the

NMR structure of *Hp*TonB(194–285), which was constructed based on the sequence homology with other TonB proteins (Fig. 9a). The NMR structure of *Hp*TonB(194–285)(PDB entry: 5LW8) revealed the presence of an additional N-terminal helix compared with the CTD of *E. coli* TonB(152–239)(PDB entry: 1XX3 [43]), which is distinctly different from any other previously reported TonB proteins but more similar to the related HasB protein [42–44]. The comparison of the [^{1}H, ^{15}N]-HSQC spectra of the truncated CTD of *Hp*TonB(194–285) and the segmentally labeled *Hp*TonB(36–285) identified clear differences in the [^{1}H, ^{15}N]-HSQC spectra (Fig. 8). The chemical shifts differences between the two constructs of *Hp*TonB are visualized on the NMR structure of *Hp*TonB(194–285) (Fig. 9b and c). The observed large chemical shift changes are located at the N terminus as expected and are also observed for the β-strands, particularly for the C-terminal strand. This data suggests that the prolonged N-terminal region of *Hp*TonB(36–285) has structural influences on the remaining CTD. We think that the N-terminal Pro-rich region affects the dynamics of the first helix, which causes chemical shift changes and might influence the biological function because the C-terminal strand is found to interact with the TonB-dependent outer membrane transporters [43–46]. This example of segmental isotopic labeling of intact *Hp*TonB with the native sequence clearly emphasizes the importance of structural investigation in the full-length context and that truncated domains of proteins, which are often used for structural investigation, might not represent actual biological conformation.

Discussion

This study reports that inteins from extremely halophilic archaea can be halo-obligatory and are thus inactive under low salinity conditions but can be activated by increasing the salt concentration. This salt-dependent activation of halo-obligatory inteins opens new opportunities to control protein splicing. For example, a salt-inducible intein-mediated protein purification system could both prevent premature cleavages and be used for purification without any reducing agents (Fig. 3c). The use of halo-obligatory inteins for backbone cyclization facilitates the

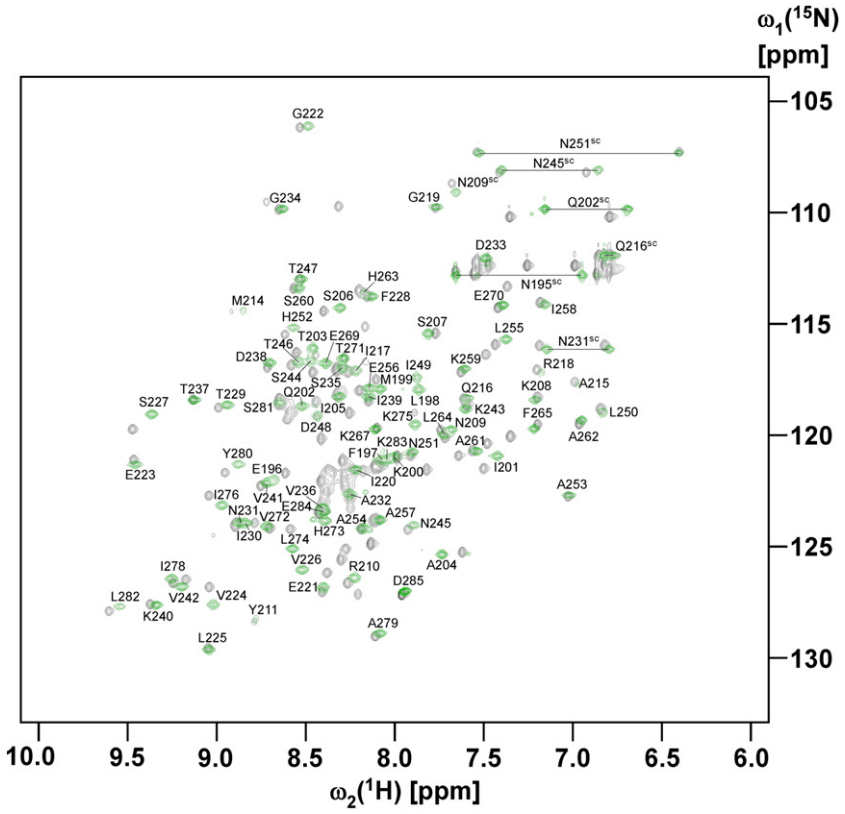

Fig. 8. An overlay of two [^{1}H, ^{15}N]-HSQC spectra of *Hp*TonB(194–285) (green) and segmentally [^{15}N, (155–285)]-labeled *Hp*TonB(36–285) (dark grey). The resonance assignments of the backbone amides of *Hp*TonB(194–285) are marked by the residue number and one-letter amino acid codes. Side-chain amide resonances of Gln and Asn are indicated by "sc".

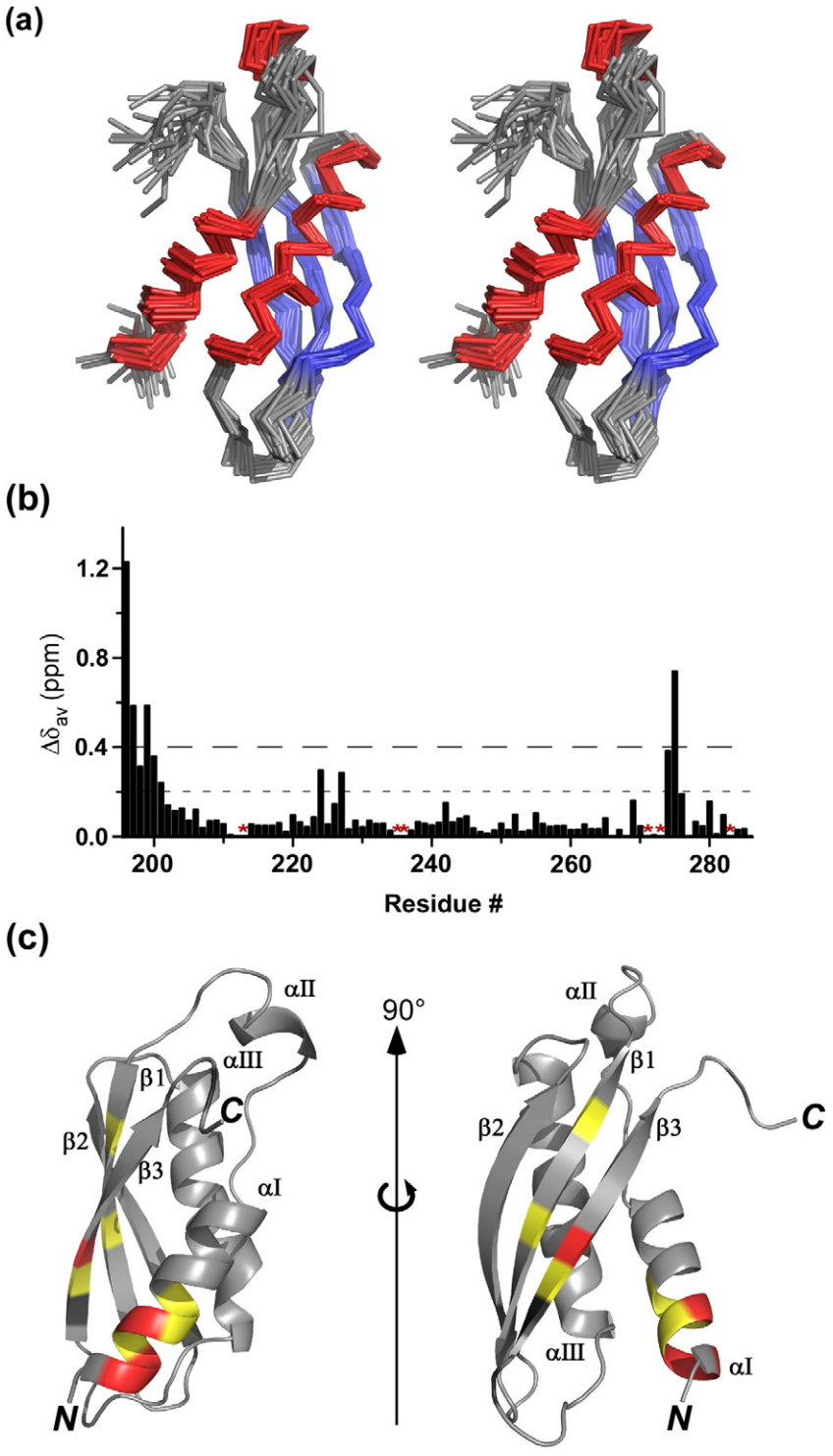

Fig. 9. (a) A stereoview of an ensemble of 20 NMR solution structures of CTD, *Hp*TonB(194–285). α-helices and β-sheets are colored in red and blue, respectively. (b) Chemical shift perturbation of *Hp*TonB(194–285) by the Pro-rich region. The average chemical differences between *Hp*TonB(194–285) and *Hp*TonB(36–285) were calculated with the formula $\Delta\delta_{av} = [(\Delta\delta_H)^2 + 0.154*(\Delta\delta_N)^2]^{1/2}$ and plotted against the residue number. (c) The locations with larger chemical shift changes ($\Delta\delta_{av} \geq 0.2$) are highlighted on the schematic drawing of NMR structure of *Hp*TonB(194–285). Residues with $\Delta\delta_{av} \geq 0.4$ ppm and 0.4 ppm $> \Delta\delta_{av} \geq 0.2$ ppm are colored in red and yellow, respectively. The N and C termini are indicated by N and C, respectively. Residues with ambiguous assignments in the segmentally [^{15}N, (155–285)]-labeled *Hp*TonB(36–285) are shown in dark gray in the structure and are indicated by red asterisk in the plot. Drawings of the structure were produced by PyMOL [62].

production of cyclic peptides and proteins *in vitro* because spontaneous cyclization *in vivo* can be prevented, thereby making it possible to incorporate purification tags in the precursor instead of the cyclized peptide (Fig. 4) [34]. This will be particularly advantageous for the backbone cyclization of smaller bioactive peptides as any tag could potentially reduce their biological activities.

These halo-obligatory inteins were found to be highly soluble both with and without a high salt concentration because the two strictly halophilic inteins could be easily concentrated to >4 mM (>80 mg/ml).

Aggregation-prone split fragments of many artificially split inteins derived from non-halophilic inteins have been problematic for wider applications of artificially split inteins despite their robust *cis*-splicing activities [24]. Artificially split inteins from halo-obligatory inteins could circumvent the solubility issue, thereby facilitating *in vitro* protein ligation using PTS [40,47]. As halo-obligatory inteins are inactive at low salt concentrations such as in the intracellular environment of *E. coli*, this system could also be combined with other split intein systems, like a naturally split DnaE intein with very fast ligation kinetics for three-fragment ligation because naturally occurring split inteins and strictly halophilic inteins will not cross-react with each other under low salinity conditions [48,49]. For example, three-fragment ligation can be achieved by two ligation steps using both *in vivo* PTS using a naturally split DnaE intein and *in vitro* PTS using an artificially split halo-obligatory intein [48,49]. This could theoretically make it possible to not only label any region of the protein but also expand the biotechnological applications of PTS for semi-synthesis in a controlled manner. For such applications, *Hut*MCM2 intein has several advantages over *Hsa* inteins and naturally split DnaE inteins. *Hut*MCM2 intein requires less salt concentration than *Hsa* inteins, which is in agreement with the growth condition of *H. utahensis* under lower salinity. The artificially split *Hut*MCM2 intein has as fast ligation kinetics as the highly efficient *Npu*DnaE intein. The *Hut*MCM2 intein uses Ser (instead of Cys, which is used by the *Npu*DnaE intein) as the nucleophilic residue at the +1 position, which will remain in the ligated product. Ser at the +1 position is more favorable for many applications such as segmental isotopic labeling, not only because it is not sensitive to oxidative environment but also because Ser is more abundant than Cys and could be exploited as the +1 position of *Hut*MCM2 intein for protein ligation. We demonstrated the practical utility of the scarless protein ligation with intact *Hp*TonB having the native sequence for NMR investigations by utilizing the newly designed split *Hut*MCM2 intein.

In conclusion, this study showed various applications of salt-inducible CPS in *cis* and *trans*. Salt-inducible CPS opens new, exciting applications as a biotechnological tool for the production of proteins and cyclic proteins/peptides by controlled ligation/cleavages. Segmental isotopic labeling of intact proteins without any sequence modification by the novel and salt-inducible split intein could facilitate the NMR structural investigation of the protein domains in their larger, full-length contexts with intact primary structures, despite flexible regions or repeating regions [48]. Protein engineering of halo-obligatory inteins or other halo-obligatory enzymes could also widen the applications of salt-inducible conditional splicing/activation as a novel protein-engineering tool.

Materials and Methods

Plasmid constructions

For comparison of protein splicing of inteins from haloarchaea

The previously described *cis*-splicing vectors of CDC21 intein (pSKDuet24) and PolII intein (pSKDuet25) from *H. salinarum* were used [28]. MCM2 intein from *H. utahensis* was cloned from genomic DNA (DSM-12940) using two oligonucleotides, I119: 5′-TTGGTACCTCCGGCATTATGCACCAATAC and I120: 5′- TAGGTACCGTCCTCGGAATTATGGACGACCATTCC, and was cloned between the *Bam*HI and *Kpn*I sites in pSKDuet24, resulting in the plasmid pSADuet616 (H_6-GB1-*Hut*MCM2-GB1).

For salt-induced protein cyclization

The C-terminal fragment (residues 121–182) of *Hsa*CDC21 intein bearing the N-terminal His-tag was produced by PCR from pSKDuet24 as the template using the three oligonucleotides of I020: 5′-CCATCACCACACTAGTGGGCAGGTCGCACCCGACG, I021: 5′-ACATATGGGCAGCCATCATCACCATCACCACACTAGTG, and I019: 5′-ATCTAGAACCCATCTGGGAGTTGTGCGAGAC. The PCR product contains *Nde*I and *Xba*I. The gene of GFP, together with a thrombin cleavage site, was amplified from pIWT63-80 [34] using the two oligonucleotides of 86: 5′- TAGGTACCGCGTGGCACCAACCCAGCAGCWGTTAC and 149: 5′-CGGGATCCAACGAGCCGAGGACGTTC. The gene for the N-terminal fragment (residues 1–120) was amplified using two oligonucleotides of I022: 5′-TCAAAGCTTAGGAGCCGTCGCCGTCCGGG and I018: 5′-TTGGTACCGGCAAGTGCGTGCGGGGCGAC from pSKDuet24. The three gene fragments were assembled into a pBAD vector (Invitrogen) using the restriction sites of *Nde*I, *Xba*I, *Kpn*I, and *Hind*III. The resulted plasmid pSABAD570 encodes the precursor protein of H_6-*Hsa*CDC21-$_{C62}$-GFP-*Hsa*CDC21$_{\Delta C62}$ under an arabinose promoter.

For salt-inducible PTS

The gene of the N-terminal split intein fragment of *Hut*MCM2 intein (*Hut*MCM2$_{\Delta C62}$) was amplified from pSADuet616 using the two oligonucleotides, I119: 5′-TCGGATCCATGCGGTGCGTTACTGGGGATACTCTC and I140: 5′- TCAAGCTTAACCGTCAGTTGCCATCGCTG, and was cloned into pSKDuet1 [18] using *Bam*HI and *Hind*III sites, resulting in pSADuet620. The gene of the C-terminal split intein fragment of *Hut*MCM2 intein (*Hut*MCM2$_{C42}$) was obtained by PCR from pSADuet616 using two oligonucleotides, I139: 5′-GTCATATGGGCGATATCGGGCTTCGA and I120: 5′-TAGGTACCGTCCTCGGAATTATGGACGACCATTCC, and was digested by *Nde*I and *Kpn*I. The gene of GB1-H_6 was obtained from pMHBAD14 (Addgene #42304) [50] by digesting with *Kpn*I and *Hind*III. The digested genes of *Hut*MCM2$_{C42}$ and GB1-H_6 were cloned into pHYRSF-1 (Addgene #34549) [40] in a stepwise manner using *Nde*I, *Kpn*I, and *Hind*III sites, which resulted in pSARSF619 bearing the gene of H_6-*Hut*MCM2$_{C42}$-GB1-H_6.

N-terminal precursor, HpTonB(36–154)-HutMCM$_{\Delta C62}$-H$_8$ for segmental labeling of HpTonB

The gene of the N-terminal half of *Hut*MCM2 intein with chitin binding domain (CBD) and octa-histidine tag (*Hut*MCM$_{\Delta C62}$-CBD-H$_8$) was purchased from IDT DNA as the cloned plasmid in pIDTSMART-KAN. The CBD domain was removed by inverse PCR using the two oligonucleotides of I731: 5′-CCGGCGGAATCCGGAGGAGAATT CGGTCACCATCATC and I732: ATGGTGACCGAA TTCTCCTCCGGATTCCGCCGGTCCTTC. The gene of *Hp*TonB(36–154) was first amplified from genomic DNA of *H. pylori* (Marshall et al., ATCC:700392D-5) using the two oligonucleotides, I816: 5′CTCATATGCGCGAAGACG CCCCAGAGCCTTTAG and I822: 5′-GTATCCCCAGTAA CGCATTTTTCTTTAGCTTCCTCTTTAG, followed by another PCR amplification with the two oligonucleotides, I816 and I646: 5′-TGCACTAGTGTATCCCCAGTAACGCANC. The amplified gene and the gene of *Hp*TonB(36–154) were cloned into a vector pHYDuet194 in a stepwise manner using the restriction sites of *Nde*I, *Bam*HI, *Hind*III, and *Spe*I. The plasmid pHYDuet194 was derived from pHYRSF1 (Addgene #34549) [40] and bears a mutation of *Nco*I-to-*Nde*I site at the start codon. The resulting plasmid pJTRSF9 contains Arg154 instead of Lys154 at the −1 splicing junction of *Hp*TonB(36–154)-*Hut*MCM$_{\Delta C62}$-H$_8$.

The plasmid pACRSF5 bearing the wild-type TonB residue of Lys154 was constructed using pJTRSF9 as the template with the two oligonucleotides, I816 and I933: 5′-GAACTAGTGTATCCCCAGTAACGCATTTTTCTT TAGCTTC, and was cloned in the same way as pJTRSF9 using *Nde*I and *Spe*I sites.

C-terminal precursor, H$_6$-HutMCM2$_{C42}$-HpTonB (155–285) for segmental labeling of HpTonB

The gene of *Hp*TonB (residues 155–285) was amplified from the genomic DNA of *H. pylori* using two oligonucleotides, I849: 5′-GGAATGGTCGTCCATAATAGCGCTCCT AAACAAGTAACAAC and I850: 5′-CAGCGGTTTCTTTAC CAAAGCTTAGTCTTCTTTCAAGCTATAAGC. The amplified DNA fragment was used for overlap extension PCR cloning [51] using plasmid pSARSF619 as the template, resulting in pBHRSF165.

For the full-length HpTonB(36–285)

For the production of the full-length *Hp*TonB without TM, the gene of TonB was amplified from the genomic DNA of *H. pylori* using the following two oligonucleotides, I816: 5′-CTCATATGCGCGAAGACGCCCCAGAGCCTTTAG and I915: 5′- GTGGTACCTTAGTCTTCTTTCAAGCTATAAGC. The amplified PCR product was cloned into pHYRSF53 [52] using *Bam*HI and *Hind*III, resulting in pACRSF5 coding for H$_6$-Smt3-*Hp*TonB(36–285) fusion protein.

Salt-induced CPS in *cis*

Small- and large-scale expression was done using plasmids pSKDuet24, pSKDuet25, or pSADuet616 in *E. coli* ER2566 cells (New England Biolab), either in 5 ml or 2-L LB medium for large-scale expression, was supplemented with 25 μg/mL kanamycin, and was grown at 37 °C. The cell culture was induced with IPTG when OD$_{600}$ reached 0.6 and was grown for another 4 h before it was harvested by centrifugation. Bacterial cells from the small-scale expression were lysed using B-PER® Bacterial Protein Extraction Reagent (ThermoFischer Scientific) and purified using Ni-NTA spin columns (GE Healthcare). The protein was eluted with elution buffer [300 mM NaCl, 250 mM Imidazole, and 50 mM sodium phosphate (pH 8.0); Buffer B]. The harvested cells from 2-L culture were lysed with EmulsiFlex C3 cell homogenizer for 10 min at 15,000 PSI and were centrifuged at 37,000*g* for 1 h at 4 °C. The cleared supernatant was loaded onto a HP HisTrap column (GE Healthcare) with loading buffer A [300 mM NaCl and 50 mM sodium phosphate (pH 8.0); Buffer A]. The protein was eluted from the column with Buffer B and dialyzed against Milli-Q water before it was concentrated. The estimated amount of the purified protein was >30 mg/L for each of the three proteins. The proteins were either tested for protein splicing immediately or stored at −70 °C.

Cis-splicing of the purified *Hsa*CDC21, *Hsa*PolII, and *Hut*MCM2 inteins was tested in either 4 M NaCl or various NaCl concentrations (1, 2, 3, and 4 M). A final concentration of 50 μM protein was incubated in a final concentration of 50 mM Tris–HCl and 4 M NaCl (pH 7.0) in the presence of 0.5 mM TCEP. The reaction mixtures were incubated for 18 h (overnight incubation) at 25 °C and stopped by mixing with 1× SDS sample buffer. The samples were analyzed on 18% SDS polyacrylamide gels.

For the time course of *cis*-splicing and kinetics at different NaCl concentrations, 100 μM protein solutions of the purified *cis*-splicing precursors containing *Hsa*CDC21 or *Hut*MCM2 inteins were induced in the presence of 0.5 mM TCEP with a final concentration of 4 M or 3.5 M NaCl, respectively. As a control, the same protein solutions were also incubated without any NaCl. The reaction mixtures were incubated for 22 h at 37 °C. For the kinetic analysis, samples were collected at given time points for SDS-PAGE analysis (0–1260 min). The reactions were stopped by mixing with SDS sample loading buffer and were analyzed on 18% SDS-PAGE, followed by staining with PhastGel Blue R (GE Healthcare). Intensities of the bands corresponding to the precursors and products were quantified from Coomassie-stained SDS-PAGE using the software Image J (NIH, Bethesda, MD, USA) and were normalized according to their molecule sizes. The ligation efficiency was estimated from the amount of the ligation product against the sum of the unreacted precursor and the side-products produced by the cleavages. Kinetic parameters of the ligation reaction were obtained by fitting the first-order kinetics function to the band intensities without any offset using SigmaPlot (Systat Software Inc). All of the experiments were performed in triplicates. The lower *cis*-splicing efficiency of *Hsa*CDC21 intein is due to the competing cleavage reactions.

Protein purification by salt-induced cleavage

GFP and GB1 were used as target proteins in a model system to create a fusion protein bearing *Hut*MCM2 with the double mutations of Ala at the C-terminal residue and the +1 position, followed by a CBD with the C-terminal octa-histidine tag. GFP–intein fusion was cloned into a pET vector, resulting in pBH(etGFP)Syn13, and GB1–intein fusion into a pDuet vector, resulting in pSADuet735. These fusion proteins were expressed in the *E. coli* ER2566 and

purified by IMAC. The cleavage of the fusion proteins was induced at a final concentration of 4 M NaCl at room temperature with 50 mM $NH_4(OH)$ or at a final concentration of 4 M NaCl without reducing agent or with either 0.5 mM TCEP or 30 mM DTT at 37 °C. The reaction mixture was purified by either a Ni-NTA column or a chitin column in order to remove the intein tag and unreacted precursor.

Salt-induced protein cyclization of GFP

The fusion protein bearing GFP and the split *Hsa*CDC21 intein was expressed in the *E. coli* ER2566 strain using the plasmid pSABAD570. The cells were grown in 0.5-L LB media supplemented with ampicillin (100 µg/mL) at 37 °C and induced with a final concentration of 0.08% arabinose and induced for another 6 h at 30 °C. The precursor protein was purified by IMAC using a HisTrap column, dialyzed against Milli-Q water at 4 °C overnight, and concentrated using a centrifugal concentrator device.

The salt-inducible cyclization was performed in two ways (rapid dilution and dialysis). For rapid dilution, the precursor solution containing no salt was mixed with a NaCl solution at a final concentration of 100 µM precursor and 3.5 M NaCl in the presence of 0.5 mM TCEP and was incubated at 37 °C for 18 h. For the induction of protein cyclization by dialysis, the protein was mixed with NaCl at a final concentration of 1.2 M NaCl, followed by dialysis against 2 L of a 3.5 M NaCl solution at 37 °C overnight (for 18 h). The solution, which was dialyzed against a 3.5M NaCl solution, was diluted to ~1 M NaCl concentration and subsequently loaded on a 5-mL HisTrap HP column (GE Healthcare) to remove the unreacted precursor. The flow-through fractions containing the cyclized GFP from the second IMAC purification were concentrated and exchanged with PBS buffer using a centrifugal concentrator device MWCO 3500 (Millipore). To confirm the backbone cyclization of GFP, we added 0.25 U of thrombin (Roche Diagnostics GmbH) to 60 µM of cyclized GFP in the presence of 0.5 mM TCEP at 37 °C, and it was analyzed by SDS-PAGE.

Conditional PTS using split *Hut*MCM2 intein

Two split precursors of $HutMCM2_{C42}$-GB1-H_6 (pSARSF619) and H_6-GB1-$HutMCM2_{\Delta C62}$ (pSADuet620) were purified by IMAC using the hexahistidine tag. The two precursors were mixed in an equimolar concentration of 30 µM and incubated in 4 M NaCl in the presence of 0.5 mM TCEP at 37 °C for 22 h. Samples were analyzed by SDS-PAGE, and the ligated product was confirmed by Matrix Assisted Laser Desorption/Ionization - Time of Flight (MALDI-TOF) mass spectrometry.

Solubility comparison of N-terminal precursors containing split inteins

The solubility of the N-terminal precursor proteins using an identical extein of GB1 was done by the comparison of soluble and insoluble fractions after the cell lysis. The plasmids for H_6-GB1-$SspDnaE_N$ (pJJDuet30) [17], H_6-GB1-$NpuDnaE_N$ (pSKDuet1), and H_6-GB1-$HutMCM2_N$ (aforementioned pSADuet620) were used for the comparison. Expression was done using *E. coli* ER2566 cells in 5-ml LB medium supplemented with 25 µg/ml kanamycin at 37 °C. Expression was induced with a final concentration of 1 mM IPTG when OD_{600} reached 0.6. The cells were grown for an additional 4 h at 37 °C before harvesting by centrifugation. The cells were lysed using B-PER® Bacterial Protein Extraction Reagent. The soluble fraction was separated from the pellet by centrifugation at 21,000*g* for 5 min. The pellet was resuspended in 1× SDS sample loading buffer and analyzed in comparison with the soluble fraction on 18% SDS PAGE.

Segmental isotopic labeling of *Hp*TonB

Expression and purification

^{15}N-labeled C-terminal precursor and the uniformly ^{15}N-labeled full-length protein were produced using plasmid pBHRSF165 or pACRSF4 in *E. coli* ER2566 strain in 2-L M9-medium at 37 °C supplemented with 25 µg/ml kanamycin containing $^{15}NH_4Cl$ as the sole nitrogen source. Protein expression was induced by the addition of a final concentration of 1 mM IPTG when OD_{600} reached 0.6, followed by an additional 5-h incubation at 30 °C before the cells were harvested by centrifugation. Expression of the unlabeled N-terminal precursor, using plasmid pACRSF5 or pJTRSF9, was done in *E. coli* ER2566 strain. The cells were grown in 2-L LB medium supplemented with 25 µg/ml kanamycin at 37 °C. When OD_{600} reached 0.6, the protein expression was induced with a final concentration of 1 mM IPTG, followed by another 4-h incubation before the cells were harvested by centrifugation.

For purification, the cells were resuspended in Buffer A and lysed using EmulsiFlex C3 homogenizer before centrifugation at 37,000*g* for 1 h at 4 °C. IMAC purification was done using HisTrap HP column (GE Healthcare) and loading Buffer A. After elution with Buffer B, the protein solution was dialyzed against PBS buffer, followed by concentration to 0.2–0.6 mM. About 20 mg/L and 30 mg/L of the precursors were purified for the ^{15}N-labeled C-terminal precursor using pBHRSF165 and for the unlabeled N-terminal precursors using either pACRSF5 or pJTRSF9, respectively (Supplementary Fig. 3b).

The fusion protein of H_6-Smt3-*Hp*TonB(36–285)(pACRSF4) was purified by IMAC, followed by Ulp1 digestion before a second IMAC round as described previously [52]. The purified *Hp*TonB(36–285) protein was dialyzed against 20 mM NaPi, pH 6.0 and concentrated to 0.5 mM for NMR studies.

Protein ligation by salt-induced PTS

The purified ^{15}N-labeled C-terminal precursor and the unlabeled N-terminal precursor were mixed at a molar ratio of 1:1. A final concentration of ~0.1 mM (14 mg and 11 mg for the N- and the C-terminal precursor, respectively) was used for the ligation of *Hp*TonB(36–154)-$HutMCM_{\Delta C62}$-H_8 with Lys154 and H_6-$HutMCM2_{C42}$-*Hp*TonB(155–285). A final concentration of ~0.4 mM was used for *Hp*TonB(36–154)-$HutMCM_{\Delta C62}$-H_8 with Arg154 and H_6-$HutMCM2_{C42}$-*Hp*TonB(155–285). The reaction was performed in a volume of 4.5 ml and dialyzed against 1 L of 0.5 M NaPi and 3.5 M NaCl at pH 7.0 in the presence of 0.5 mM TCEP. The reaction mixture was dialyzed for 16–40 h at room temperature in

a dialysis tube with MWCO of 6–8 kDa. The ligation reaction was stopped by diluting the reaction mixture to a final concentration of 50 mM NaPi and 150 mM NaCl. For SDS-PAGE analysis, samples (20 μl) were taken from the reaction mixture after 2, 16, 22, and 39 h. The samples were diluted by mixing each sample with 1× SDS loading buffer (60 μl) and were analyzed on 18% SDS polyacrylamide gels after incubation at 95 °C for 5 min. As a control for the NaCl induction, an identical reaction mixture was dialyzed against PBS buffer instead of 3.5-M NaCl solution. The samples for SDS-PAGE analysis were taken at the same time points. The ligation mixture after dialysis was further purified by IMAC using a HisTrap HP column to remove the unreacted precursors and excised intein fragments. The flow-through fractions from IMAC were pooled and dialyzed against 20 mM NaPi at pH 6.0 for further analysis by NMR and mass spectrometry. The total amount of segmentally ^{15}N-labeled ligated product after the second purification was about 3 mg for the segmentally [^{15}N, (155–285)]-labeled *Hp*TonB(36–285) when the plasmids of pACRSF5 and pBHRSF165 were used for the expression of the precursors. Then, 0.2-mM (ligated product) solutions of the segmentally [^{15}N, (155–285)]-labeled *Hp*TonB(36–285) were prepared for NMR measurements in a 250-μl volume of 20 mM NaPi, pH 6.0 containing 10% D_2O.

NMR measurements

NMR measurements for the *Hp*TonB samples (20 mM NaPi, pH 6.0 containing 10% D_2O) in shigemi tubes (250 μl) were recorded at the ^{1}H frequency of 850 MHz on a Bruker Avance III HD spectrometer equipped with a triple-resonance cryogenic probe at 298 K.

NMR solution structure of *Hp*TonB(194–285)

Doubly [^{15}N, ^{13}C]-labeled protein sample of *Hp*TonB (194–285) for the NMR structure determination was expressed in *E. coli* ER2566, which is transformed with the plasmid pACRSF01 [52] in 2-L M9 medium supplemented with kanamycin (25 μg/ml), containing $^{15}NH_4Cl$ and ^{13}C-glucose as sole nitrogen and carbon source, respectively, and purified as described previously [52]. The protein was dialyzed against 20 mM NaPi, pH 6.0 and was concentrated to 0.5 mM.

For sequential backbone assignment, the following experiments were used: [^{1}H, ^{15}N]-HSQC, HNCO, HNCA, HNCACB, HN(CO)CA, HN(CA)CO, and CBCA(CO)NH [53]. Aliphatic side chains were assigned using [^{1}H, ^{13}C]-HSQC, HCCH-correlated spectroscopy(COSY), CC(CO)NH, H(CCO)NH, HBHA(CO)NH, and ^{15}N-edited [^{1}H, ^{1}H]-total correlated spectroscopy (TOCSY) experiments, and aromatic side chain assignment is based on ct-[^{1}H, ^{13}C]-HSQC and ^{13}C-edited [^{1}H, ^{1}H]-NOESY-HSQC spectra. Chemical shift assignment was performed using CcpNmr Analysis software [54]. Solution structure conformers were generated using CYANA 3.0 software, based on automated NOESY cross peaks assignment [55,56]. Empirical dihedral angle constraints, based on assigned chemical shifts, were generated with TALOS software [57]. NOE distance constraints were obtained from 3D ^{15}N- and ^{13}C-edited [^{1}H, ^{1}H]-NOESY-HSQC spectra with mixing times of 90 ms. Restrained energy minimization was performed using the 20 best conformers with the lowest CYANA target function by AMBER 14 [58,59]. The structures were validated using PSVS 1.5 [60] and WHAT IF [61]. Structural statistics are shown in Supplementary Table 1.

Accession numbers

The atomic coordinates and assigned chemical shifts have been deposited in the Protein Data Bank† (PDB ID: 5LW8; BMRB entry: 34043).

Acknowledgments

The authors thank the Protein Chemistry Research Group for mass spectrometry; J. Tommila, S. Ferkau, and B. Haas for their technical help; and J.S. Oeemig for his assistance. This work is supported by the grants from the Academy of Finland (118385, 1131413, 1137995, and 1277335) and by the Sigrid Jusélius Foundation. The NMR facility is supported by Biocenter Finland. A.C. was supported by the National Doctoral Programme in Informational and Structural Biology. CSC–IT Center for Science Ltd. is acknowledged for the allocation of computational resources.

Competing financial interests: The authors declare no competing financial interests.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2016.10.006.

Received 23 July 2016;
Received in revised form 29 September 2016;
Accepted 1 October 2016
Available online 6 October 2016

Keywords:
inteins;
halophiles;
protein ligation;
NMR spectroscopy;
segmental labeling

Present address: A. S. Aranko, School of Chemical Technology, Aalto University, P.O. Box 16100, FI–00076, Espoo, Finland.
Present address: I. Tascon, Institute of Biochemistry, Biocenter, Goethe University Frankfurt, Max-von-Laue-Str. 9, D-Frankfurt/Main, Germany.
†www.pdb.org

Abbreviations used:
PTS, protein *trans*-splicing; GB1, the B1 domain of IgG binding protein G; HSQC, heteronuclear single quantum coherence; TM, transmembrane; CTD, C-terminal domain; IMAC, immobilized metal chelating; TCEP, tris(2-carboxyethyl)phosphine; CPS, conditional protein splicing; GFP, green fluorescent protein; CBD, chitin binding domain; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; PBS, Phosphate-buffered saline.

References

[1] A. Oren, in: M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer, E. Stackebrandt (Eds.), Life at high salt concentrations. The Prokaryotes. A handbook of the Biology of Bacteria: Ecophysiology and Biochemistry, 2, Springer, New York 2006, pp. 263–282.

[2] R. Jaenicke, Protein stability and molecular adaptation to extreme conditions, Eur. J. Biochem. 202 (1991) 715–728.

[3] D. Madern, C. Ebel, G. Zaccai, Halophilic adaptation of enzymes, Extremophiles 4 (2000) 91–98.

[4] F.B. Perler, InBase, the intein database, Nucleic Acids Res. 30 (2002) 383–384.

[5] S.M. Soucy, M.S. Fullmer, R.T. Papke, J.P. Gogarten, Inteins as indicators of gene flow in the halobacteria, Front. Microbiol. 5 (2014) 299.

[6] R. Hirata, Y. Ohsumk, A. Nakano, H. Kawasaki, K. Suzuki, Y. Anraku, Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*, J. Biol. Chem. 265 (1990) 6726–6733.

[7] H. Paulus, Protein splicing and related forms of protein autoprocessing, Annu. Rev. Biochem. 69 (2000) 447–496.

[8] N. Sakakibara, M. Han, C.R. Rollor, R.C. Gilson, C. Busch, G. Heo, Z. Kelman, Cloning, purification, and partial characterization of the *Halobacterium sp.* NRC-1 minichromosome maintenance (MCM) helicase, Open Microbiol. J. 2 (2008) 13–17.

[9] B.P. Callahan, N.I. Topilina, M.J. Stanger, P. Van Roey, M. Belfort, Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications, Nat. Struct. Mol. Biol. 18 (2011) 630–633.

[10] N.I. Topilina, O. Novikova, M. Stanger, N.K. Banavali, M. Belfort, Post-translational environmental switch of RadA activity by extein–intein interactions in protein splicing, Nucleic Acids Res. 43 (2015) 6631–6648.

[11] H. Mootz, T. Muir, Protein splicing triggered by a small molecule, J. Am. Chem. Soc. 124 (2002) 9044–9045.

[12] M.P. Zeidler, C. Tan, Y. Bellaiche, S. Cherry, S. Häder, U. Gayko, N. Perrimon, Temperature-sensitive control of protein activity by conditionally splicing inteins, Nat. Biotechnol. 22 (2004) 871–876.

[13] A.S. Aranko, J.S. Oeemig, T. Kajander, H. Iwaï, Intermolecular domain swapping induces intein-mediated protein alternative splicing, Nat. Chem. Biol. 9 (2013) 616–622.

[14] S. Wong, A.A. Mosabbir, K. Truong, An engineered split intein for photoactivated protein *trans*-splicing, PLoS One 10 (2015), e0135965.

[15] M.W. Southworth, E. Adam, D. Panne, R. Byer, R. Kautz, F.B. Perler, Control of protein splicing by intein fragment reassembly, EMBO J. 17 (1998) 918–926.

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

[17] S. Züger, H. Iwai, Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies, Nat. Biotechnol. 23 (2005) 736–740.

[18] H. Iwai, S. Züger, J. Jin, P.H. Tam, Highly efficient protein *trans*-splicing by a naturally split DnaE intein from *Nostoc punctiforme*, FEBS Lett. 580 (2006) 1853–1858.

[19] A.S. Aranko, S. Züger, E. Buchinger, H. Iwaï, *In vivo* and *in vitro* protein ligation by naturally occurring and engineered split DnaE inteins, PLoS One 4 (2009), e5185.

[20] J.H. Appleby-Tagoe, I.V. Thiel, Y. Wang, Y. Wang, H.D. Mootz, X.Q. Liu, Highly efficient and more general *cis*- and *trans*-splicing inteins through sequential directed evolution, J. Biol. Chem. 286 (2011) 34,440–34,447.

[21] P. Carvajal-Vallejos, R. Pallissé, H.D. Mootz, S.R. Schmidt, Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources, J. Biol. Chem. 287 (2012) 28,686–28,696.

[22] M. Cheriyan, C.S. Pedamallu, K. Tori, F. Perler, Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues, J. Biol. Chem. 288 (2013) 6202–6211.

[23] N.H. Shah, G.P. Dann, M. Vila-Perelló, Z. Liu, T.W. Muir, Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering, J. Am. Chem. Soc. 134 (2012) 11,338–11,341.

[24] A.S. Aranko, J.S. Oeemig, D. Zhou, T. Kajander, A. Wlodawer, H. Iwaï, Structure-based engineering and comparison of novel split inteins for protein ligation, Mol. BioSyst. 10 (2014) 1023–1034.

[25] A.S. Aranko, A. Wlodawer, H. Iwaï, Nature's recipe for splitting inteins, Protein Eng. Des. Sel. 27 (2014) 263–271.

[26] G. Volkmann, X.Q. Liu, Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein, PLoS One 4 (2009), e8381.

[27] H. Wu, Z. Hu, X.Q. Liu, Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 9226–9231.

[28] S. Ellilä, J.M. Jurvansuu, H. Iwaï, Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in *Escherichia coli*, FEBS Lett. 585 (2011) 3471–3477.

[29] J.N. Reitter, C.E. Cousin, M.C. Nicastri, M.V. Jaramillo, K.V. Mills, Salt-dependent conditional protein splicing of an intein from *Halobacterium salinarum*, Biochemistry 55 (2016) 1279–1282.

[30] C.J. Noren, J. Wang, F.B. Perler, Dissecting the chemistry of protein splicing and its applications, Angew. Chem. Int. Ed. Engl. 39 (2000) 450–466.

[31] S. Chong, F.B. Mersha, D.G. Comb, M.E. Scott, D. Landry, L.M. Vence, F.B. Perler, J. Benner, R.B. Kucera, C.A. Hirvonen, J.J. Pelletier, H. Paulus, M.Q. Xu, Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element, Gene 192 (1997) 271–281.

[32] M.W. Southworth, K. Amaya, T.C. Evans, M.Q. Xu, F.B. Perler, Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein, BioTechniques 27 (1999) 118–120.

[33] Y. Minato, T. Ueda, A. Machiyama, I. Shimada, H. Iwaï, 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 (2012) 191–207.

[34] H. Iwai, A. Lingel, A. Plückthun, Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-PfuI intein from *Pyrococcus furiosus*, J. Biol. Chem. 276 (2001) 16,548–16,554.
[35] H. Iwai, A. Plückthun, Circular beta-lactamase: stability enhancement by cyclizing the backbone, FEBS Lett. 459 (1999) 166–172.
[36] C.P. Scott, E. Abel-Santos, M. Wall, D.C. Wahnon, S.J. Benkovic, Production of cyclic peptides and proteins *in vivo*, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 13,638–13,643.
[37] S.R. Trevino, J.M. Scholtz, C.N. Pace, Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa, J. Mol. Biol. 366 (2007) 449–460.
[38] R.J. Clark, H. Fischer, L. Dempster, N.L. Daly, K.J. Rosengren, S.T. Nevin, F.A. Meunier, D.J. Adams, D.J. Craik, Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 13,767–13,772.
[39] H. Tokunaga, T. Arakawa, M. Tokunaga, Novel soluble expression technologies derived from unique properties of halophilic proteins, Appl. Microbiol. Biotechnol. 88 (2010) 1223–1231.
[40] M. Muona, A.S. Aranko, V. Raulinaitis, H. Iwaï, Segmental isotopic labeling of multi-domain and fusion proteins by protein *trans*-splicing *in vivo* and *in vitro*, Nat. Protoc. 5 (2010) 574–587.
[41] C. Chang, A. Mooser, A. Plückthun, A. Wlodawer, Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold, J. Biol. Chem. 276 (2001) 27,535–27,540.
[42] J. Ködding, F. Killig, P. Polzer, S.P. Howard, K. Diederichs, W. Welte, Crystal structure of a 92-residue C-terminal fragment of TonB from *Escherichia coli* reveals significant conformational changes compared to structures of smaller TonB fragments, J. Biol. Chem. 280 (2005) 3022–3028.
[43] S.R. Peacock, A.M. Weljie, P.S. Howard, F.D. Price, H.J. Vogel, The solution structure of the C-terminal domain of TonB and interaction studies with TonB box peptides, J. Mol. Biol. 345 (2005) 1185–1197.
[44] C.G. De Amorim, A. Prochnicka-Chalufour, P. Delepelaire, J. Lefèvre, C. Simenel, C. Wandersman, et al., The structure of HasB reveals a new class of TonB protein fold, PLoS One 8 (2013), e58964.
[45] P.D. Pawelek, N. Croteau, C. Ng-Thow-Hing, C.M. Khursigara, N. Moiseeva, J.W. Coulton, Structure of TonB in complex with FhuA, *E. coli* outer membrane receptor, Science 312 (2006) 1399–1402.
[46] N. Cadieux, R.J. Kadner, Site-directed disulfide bonding reveals an interaction site between energy-coupling protein TonB and BtuB, the outer membrane cobalamin transporter, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 10,673–10,678.
[47] T. Otomo, K. Teruya, K. Uegaki, T. Yamazaki, Y. Kyogoku, Improved segmental isotope labeling of proteins and application to a larger protein, J. Biomol. NMR 14 (1999) 105–114.
[48] A.E.L. Busche, A.S. Aranko, M. Talebzadeh-Farooji, F. Bernhard, V. Dötsch, H. Iwaï, Segmental isotopic labeling of a central domain in a multidomain protein by protein *trans*-splicing using only one robust DnaE intein, Angew. Chem. Int. Ed. Engl. 48 (2009) 6128–6131.
[49] J. Shi, T.W. Muir, Development of a tandem protein *trans*-splicing system based on native and engineered split inteins, J. Am. Chem. Soc. 127 (2005) 6198–6206.
[50] J.S. Oeemig, A.S. Aranko, J. Djupsjöbacka, K. Heinämäki, H. Iwaï, Solution structure of DnaE intein from *Nostoc punctiforme*: structural basis for the design of a new split intein suitable for site-specific chemical modification, FEBS Lett. 9 (2009) 1451–1456.
[51] F. Van den Ent, J. Löwe, RF cloning: a restriction-free method for inserting target genes into plasmids, J. Biochem. Biophys. Methods 67 (2006) 67–74.
[52] F. Guerrero, A. Ciragan, H. Iwaï, Tandem SUMO fusion vectors for improving soluble protein expression and purification, Protein Expr. Purif. 116 (2015) 42–49.
[53] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proeins in solution employing pulsed field gradients, Prog. NMR Spectroc. 34 (1999) 93–158.
[54] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 59 (2005) 687–696.
[55] P. Güntert, Automated NMR structure calculation with CYANA, Methods Mol. Biol. 278 (2004) 353–378.
[56] P. Güntert, C. Mumenthaler, K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, J. Mol. Biol. 273 (1997) 283–298.
[57] G. Cornilescu, F. Delaglio, A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology, J. Biomol. NMR 13 (1999) 289–302.
[58] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, Comput. Phys. Commun. 91 (1995) 1–41.
[59] W.D. Cornell, P. Cieplack, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[60] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by structural genomics consortia, Proteins 66 (2007) 778–795.
[61] W.G. Touw, C. Baakman, J. Black, T.A. te Beek, E. Krieger, R.P. Joosten, G. Vriend, A series of PDB-related databanks for everyday needs, Nucleic Acids Res. 43 (Database issue) (2015) D364–D368.
[62] The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.