Controllable protein cleavages through intein fragment complementation

Gerrit Volkmann,[†] Wenchang Sun,[†] and Xiang-Qin Liu*

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 1X5

Received 5 February 2009; Revised 8 September 2009; Accepted 9 September 2009

DOI: 10.1002/pro.249

Published online 18 September 2009 proteinscience.org

Abstract: Intein-based protein cleavages, if carried out in a controllable way, can be useful tools of recombinant protein purification, ligation, and cyclization. However, existing methods using contiguous inteins were often complicated by spontaneous cleavages, which could severely reduce the yield of the desired protein product. Here we demonstrate a new method of controllable cleavages without any spontaneous cleavage, using an artificial S1 split-intein consisting of an 11-aa N-intein (I_N) and a 144-aa C-intein (I_C). In a C-cleavage design, the I_C sequence was embedded in a recombinant precursor protein, and the small I_N was used as a synthetic peptide to trigger a cleavage at the C-terminus of I_C . In an N-cleavage design, the short I_N sequence was embedded in a recombinant precursor protein, and the separately produced I_C protein was used to catalyze a cleavage at the N-terminus of I_N . These N- and C-cleavages showed >95% efficiency, and both successfully avoided any spontaneous cleavage during expression and purification of the precursor proteins. The N-cleavage design also revealed an unexpected and interesting structural flexibility of the I_C protein. These findings significantly expand the effectiveness of intein-based protein cleavages, and they also reveal important insights of intein structural flexibility and fragment complementation.

Keywords: intein; split-intein; protein splicing; controllable cleavage; protein purification

Introduction

Inteins are internal protein sequences embedded in host proteins, and intein-catalyzed protein splicing excises the intein and concomitantly ligates the flanking sequences (exteins) with a peptide bond.^{1,2} Inteins catalyze protein splicing using a four-step mechanism.^{3,4} In Step 1, an N-S or N-O acyl shift breaks the peptide bond between the N-terminal extein (N-extein) and the intein, and the N-extein is linked to the side chain of the intein's first residue through an ester bond. In Step 2, a transesterification links the N-extein to the side chain of the C-extein's first residue through an ester bond. In Step 3, cyclization of the intein's last

residue (usually Asn) breaks the peptide bond between the intein and the C-extein. In Step 4, an S-N or O-N acyl shift changes the ester bond between the two exteins to a peptide bond.

Site-specific cleavages at the N- or C-terminus of an intein can result from alterations of the protein splicing reaction.³ To achieve N-terminal cleavage, the intein's last residue is mutated to abolish Step 3 (consequently also Step 4) of the splicing mechanism. Step 1 (with or without Step 2) of the splicing mechanism can still happen, and the resulting ester bond can spontaneously hydrolyze to separate the N-extein from the intein. To achieve C-cleavage, the intein's first residue is mutated to abolish Step 1 (consequently also Steps 2 and 4) of the splicing mechanism, but Step 3 (Asn cyclization) can still occur to break the peptide bond between the C-extein and the intein.

Intein-based controllable cleavages have been adapted to an increasing number of useful applications. These include the IMPACT method for single-step purification of recombinant proteins,^{5–7} the expressed protein ligation (EPL) method for joining

[†]Gerrit Volkmann and Wenchang Sun contributed equally to this work.

Grant sponsor: National Science and Engineering Research Council of Canada.

*Correspondence to: Xiang-Qin Liu, PhD Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7. E-mail: Paul.Liu@Dal.Ca

polypeptides,^{8,9} and the cyclization of recombinant proteins or peptides.¹⁰ These intein-based methods promise a wide range of practical applications,^{2,11} because inteins can generally function in a non-native extein context. However, it remains a challenge to achieve completely controllable cleavages, in which the cleavage reaction is induced only when needed, i.e., after the precursor protein has been expressed and purified.

Previously reported methods employing contiguous inteins have often encountered unwanted spontaneous cleavages. Among these methods, N-cleavage at the N-terminus of an intein was induced by a strong nucleophile such as dithiothreitol (DTT), and C-cleavage at the C-terminus of an intein was induced by a temperature shift.^{5,9,12} However, these methods were often complicated by various levels of spontaneous cleavages during protein expression and purification, which can significantly decrease the yield of the desired protein product and complicate subsequent uses of the cleaved protein. A rationally designed single mutation in an intein has been used to prevent spontaneous N-cleavage¹³; however, its mechanism and general applicability have not been reported.

Split-inteins offer a potential solution for achieving controllable cleavages without spontaneous cleavage. A split-intein consists of two fragments named the N-terminal intein (I_N) and the C-terminal intein (I_C). Natural split-inteins have been found in cyanobacteria,^{14,15} and artificial split-inteins have been engineered from contiguous inteins.^{16,17} The I_N and I_C fragments of conventional split-inteins (natural or artificial) are typically ~110 aa and ~40 aa in length, respectively. In a more recent study,¹⁸ the *Ssp* DnaB mini-intein was split at several different sites individually to test for *trans*-splicing function. One split site near the N-terminus of the intein produced a functional split-intein consisting of an 11-aa I_N and a 144-aa I_C , which was named the S1 split-intein. Because the short I_N sequence can easily be synthesized in a peptide, this S1 split-intein has been successfully used *in vitro* to *trans*-splice fluorescent labels onto the N-terminus of recombinant proteins.¹⁹

In this report, we tested the above *Ssp* DnaB S1 split-intein for controllable cleavages of recombinant proteins. The short I_N peptide was shown to trigger a site-specific cleavage at the C-terminus of I_C embedded in a recombinant protein. The short I_N was also used as a recognition sequence embedded in a recombinant protein, and the I_C protein was used to trigger a site-specific cleavage at the N-terminus of I_N . These controllable cleavages showed high efficiencies *in vitro* and completely avoided any unwanted spontaneous cleavages. The reaction rate constants of these cleavage reactions were analyzed and compared to those of conventional inteins. Potential uses and advantages of these controllable cleavages are discussed in comparison to other cleavage methods. Structural implications derived from our cleavage systems are also discussed.

Results

I_N peptide-activated C-cleavage

The *Ssp* DnaB S1 split-intein was tested for controllable C-cleavage as shown in Figure 1(A). The recombinant precursor protein consisted of the 144-aa I_C sequence of the split-intein and five native C-extein residues, flanked by an affinity-binding domain (M, maltose-binding protein) at the N-terminus and a protein of interest (T, a chitin-binding domain fused to thioredoxin) at the C-terminus. The affinity-binding domain permitted easy purification of the precursor protein from producing cells. The 11-aa I_N (sequence: CISGDSLISLA) was added as a synthetic peptide to activate C-cleavage (through Asn cyclization) at the C-terminus of the I_C sequence. This C-cleavage would convert the precursor protein $MI_C T$ into an N-terminal fragment MI_C and a C-terminal fragment T. The precursor protein and its cleavage products could easily be identified through SDS-PAGE and Western blotting, based on each protein's predicted size and specific recognition by anti-M and anti-T antibodies.

The precursor protein $MI_C T$ was expressed in *E. coli* cells, partially purified using amylose resin, and treated with I_N to trigger C-cleavage [Fig. 1(B)]. No spontaneous cleavage or splicing product was detected during the expression and purification of the precursor protein. When the purified precursor protein was incubated with the I_N peptide, the expected cleavage products (MI_C and T) appeared, showing that the I_N peptide had triggered the expected C-cleavage reaction. The efficiency of C-cleavage was estimated as the percentage of the precursor protein that had been converted into the cleavage products at the end of the incubation, using the Coomassie-stained SDS gel for measurements. Under the *in vitro* conditions used, 70 and 90% of the precursor protein underwent C-cleavage in the absence and presence of the reducing agent DTT, respectively. DTT alone did not result in C-cleavage (data not shown).

It is well known that the flanking extein residues can have a pronounced effect on intein-mediated reactions. In a previous study of C-cleavage of the contiguous *Ssp* DnaB mini-intein,⁹ a marked decrease in C-cleavage efficiency was observed when the native +1 residue was changed from serine to other amino acids with distinct chemical properties (Ser > 90%; Glu and Met ~ 50%; Trp ~ 25%; Thr ~ 15%; Leu and Arg < 10%; Pro 0%). We therefore tested these same amino acid changes in our system to investigate which effect the amino acid at the amino terminus of the protein of interest has on the C-cleavage reaction. We observed no significant decrease in C-cleavage efficiency when the native +1 residue was changed from serine to the other tested amino acids (86–91% cleavage) except for proline (30% cleavage).

The rate constant of the I_N peptide-triggered C-cleavage reaction was determined through kinetic

Figure 1. I_N peptide-triggered C-cleavage. (A) Schematic illustration of the controllable C-cleavage using the *Ssp* DnaB S1 split-intein. The recombinant precursor protein consists of the 144-aa I_C sequence flanked by two other sequences (M and T). M stands for maltose-binding protein, which is an affinity binder for protein purification. T represents a protein of interest, which is a chitin-binding domain fused with thioredoxin. The synthetic 11-aa I_N peptide is shown with its sequence. (B) SDS-PAGE analysis of the C-cleavage reaction. Protein bands were visualized by Coomassie blue staining, by Western blot using an anti-M antibody against the maltose-binding protein, or by Western blot using an anti-T antibody against thioredoxin, as indicated. Positions of the precursor protein ($MI_C T$) and the cleavage products (MI_C and T) are marked. Predicted sizes of $MI_C T$, MI_C , and T are 77, 58.7, and 18.3 kDa, respectively. Lanes 1 and 2 are total proteins of *E. coli* cells before and after IPTG-induced expression of the precursor protein, respectively. Lane 3 shows partially purified precursor protein. Lanes 4 and 5 are the precursor protein after incubation with the I_N peptide in the absence and presence of DTT, respectively.

analysis [Fig. 2(A)]. A pseudo-first order reaction was achieved by using ~ 40 -fold molar excess of the I_N peptide relative to the precursor protein. The C-cleavage reaction was analyzed at different time points over a period of 24 h by estimating the percentage of precursor protein that had been cleaved. The rate constant (k_{obs}) was calculated to be $(1.87 \pm 0.2) \times 10^{-4} \text{ s}^{-1}$ in the absence of DTT and $(2.93 \pm 0.4) \times 10^{-4} \text{ s}^{-1}$ in the presence of DTT with maximum cleavage efficiencies of $\sim 75\%$ and $\sim 95\%$, respectively.

We reasoned that the enhancing effect of DTT on the above C-cleavage reaction might be due to DTT preventing the formation of disulfide bonds between the N-terminal Cys residues of two I_N peptides. To test this hypothesis, we replaced DTT with the reducing agent TCEP, which can prevent disulfide bond forma-

tion but cannot break ester bonds. The presence of TCEP resulted in cleavage efficiency and a rate constant similar to those seen with DTT [Fig. 2(A)], which supported the above hypothesis. In a more direct test, we changed the N-terminal Cys residue of the I_N peptide to Ser to produce the I_{N2} peptide (sequence: SISGDSLISLA). In the absence of any reducing agent, this I_{N2} peptide produced high efficiencies ($>95\%$) of C-cleavage with a rate constant that was ~ 10 -fold higher than that observed for the I_N -triggered C-cleavage reaction. The addition of TCEP increased the rate constant of I_{N2} -triggered C-cleavage only marginally from $(1.31 \pm 0.5) \times 10^{-3} \text{ s}^{-1}$ to $(2.43 \pm 0.5) \times 10^{-3} \text{ s}^{-1}$ [Fig. 2(B)].

To test the generality of the peptide-triggered C-cleavage reaction, we replaced the target protein T in $MI_C T$ precursor with four different proteins including β -galactosidase, glutathione-S-transferase (GST), red fluorescent protein (RFP), and enhanced green fluorescent protein (eGFP). All four proteins underwent peptide-triggered C-cleavage yielding the MI_C fragment and the respective target protein (Fig. 3). Peptide I_{N2} was again more efficient than peptide I_N in triggering C-cleavage.

The above C-cleavage method was also modified to produce a protein of interest having a Cys residue at its N-terminus, because an N-terminal Cys is required for certain applications such as EPL. A new precursor protein ($CI_C T'$) was produced through recombinant DNA and site-directed mutagenesis, in which the affinity-binding domain (C) at the N-terminus was a chitin-binding domain, the I_C in the middle was the same as before, and the protein of interest (T') at the C-terminus was a thioredoxin having an N-terminal Cys residue. The precursor protein $CI_C T'$, which was expressed in *E. coli* and partially purified on chitin beads, was treated with the I_{N2} peptide to trigger C-cleavage at the C-terminus of I_C (Fig. 4). The expected cleavage product, the C-terminal fragment T' , was clearly observed in SDS-PAGE based on its predicted size (12 kDa), and the identity of the T' protein, which was released from the chitin beads through C-cleavage, was further confirmed by mass spectrometry (data not shown). The elution fraction containing T' also revealed the presence of the $CI_C T'$ precursor protein, which was likely released from the chitin column during the prolonged incubation period. The presence of the first 20 amino acids of maltose-binding protein, which were included in the $CI_C T'$ protein construct to yield higher expression,⁶ may have resulted in a decreased affinity binding of the chitin-binding domain to the chitin affinity matrix.

The on-column C-cleavage reaction of $CI_C T'$ to release a target protein with an N-terminal Cys residue from an affinity matrix suggested that the peptide-triggered C-cleavage reaction could generally be useful for the purification of recombinant proteins. To investigate this further, a different precursor protein was

Figure 2. Kinetic analysis of the C-cleavage reaction. The plots show time courses of the percentage of C-cleavage after incubation with peptide I_N (A) or peptide I_{N2} (B) and reducing agents as indicated (DTT: 50 mM, TCEP: 0.1 mM). C-cleavage efficiency was calculated as the ratio of MI_C over the total ($MI_C + MI_{C'T}$), which was estimated from Coomassie-stained SDS gels through densitometry analysis. The peptides were used in ~ 40 -fold molar excess to achieve a pseudo-first order reaction. All experiments were performed in triplicate, and error bars represent standard deviations.

constructed ($CI_{C'E}$), in which the Ser+1Cys mutation in $CI_{C'T}$ was reverted back to the wild-type +1 residue, and the gene for thioredoxin was replaced with that for the eGFP (or E). We also omitted the first 20 amino acid residues of maltose-binding protein because of their proposed negative effect on the affinity binding of the chitin-binding domain to the chitin resin seen with the $CI_{C'T}$ precursor protein. This $CI_{C'E}$ protein was immobilized to chitin beads after expression in *E. coli*, and C-cleavage was induced by addition of I_{N2} peptide to the resin. The Coomassie-stained SDS gel of the elution fraction shows a major protein species migrating at ~ 27 kDa, which is the expected

molecular weight of eGFP (Fig. 5). This protein reacted strongly with anti-GFP antibodies in Western blot analysis (not shown), confirming that nearly pure eGFP was released from the chitin resin via peptide-induced C-cleavage with a yield of ~ 3 mg/L of culture. Importantly, and in contrast to the results obtained with the $CI_{C'T}$ precursor protein (Fig. 4), no contaminating uncleaved precursor protein co-eluted with the

Figure 3. Robustness of peptide-triggered C-cleavage. Precursor proteins containing maltose-binding protein (M) followed by the Ssp DnaB S1 I_C fragment and the indicated target proteins were purified from *E. coli* by amylose affinity chromatography, and incubated with peptide I_N or I_{N2} in the absence of reducing agent for 16 h at 37°C. Reactions were analyzed by SDS-PAGE and Coomassie-staining. The bands corresponding to precursor proteins are marked with squares, the MI_C fragments are marked with circles, and bands highlighted with an asterisk represent the target protein (i.e., the C-terminal C-cleavage fragment) in a given reaction. Molecular weights of protein standards are given in kDa.

Figure 4. C-cleavage to produce a protein with an N-terminal Cys residue. The C-cleavage design is similar to that of Figure 1(A), except that the precursor protein ($CI_{C'T}$) consists of the I_C sequence flanked by a chitin-binding domain (C) on the N-terminal side and a thioredoxin (T') on the C-terminal side. The T' sequence begins with a Cys residue, not a Ser residue as in Figure 1. Lanes 1 and 2 are total proteins of *E. coli* cells before and after IPTG-induced expression of the precursor protein, respectively. Lane 3 is the soluble *E. coli* protein fraction prior to incubation with chitin resin, Lane 4 represents the flow-through (proteins not bound by the resin) and Lane 5 is the wash fraction. After incubation of immobilized proteins (Lane 6) with I_N peptide for 24 h at 4°C, proteins were eluted (Lane 7). The desired cleavage product (T') has a predicted size of 12 kDa, and its position is marked. Molecular weights of protein standards are given in kDa.

Figure 5. Purification of eGFP by on-column peptide-triggered C-cleavage. $ClcE$ precursor protein was expressed and immobilized to chitin beads as described in Figure 4, and incubated with I_N2 peptide for 24 h at 4°C. The elution fraction (Lane 6) indicates release of eGFP from the chitin resin as a result of C-cleavage. Proteins remaining on the resin were stripped off with 1% SDS (Lane 7). See Figure 4 for description of the other lanes.

target protein eGFP. This confirmed that the first 20 residues of maltose-binding protein present at the N-terminus of the $ClcT'$ protein indeed decreased the affinity of the chitin-binding domain for the chitin resin, resulting in co-elution of the $ClcT'$ precursor protein with the target protein T' in Figure 4.

***I_C* protein-catalyzed N-cleavage**

Controllable N-cleavage, using the *Ssp* DnaB S1 split-intein, was tested as shown in Figure 6(A). The 11-aa I_N sequence (along with five native N-extein residues) was embedded in a precursor protein having a protein of interest (M, maltose-binding protein) at the N-terminus and a tag protein (T, thioredoxin) at the C-terminus. This recombinant precursor protein was expressed in *E. coli* and purified using amylose resin. The IcH protein consisted of the I_C sequence followed by a Ser residue and a His-tag (6 His residues). The Asn residue at the end of the I_C sequence was changed to Ala to prevent any cleavage or splicing at the C-terminus of I_C . The IcH protein was expressed in *E. coli* and purified through its His-tag using metal affinity chromatography. No spontaneous cleavage or splicing was detected during the expression and purification of the precursor protein and the IcH protein.

The purified precursor protein MI_NT and the IcH protein were mixed in a molar ratio of 1:8, the mixture was incubated at room temperature overnight for N-cleavage to occur, and resulting proteins were analyzed through SDS-PAGE and Western blotting [Fig. 6(B)]. The expected cleavage product M was readily identified after staining with Coomassie Blue, based on its predicted size of 42.5 kDa. The other cleavage product, I_NT , was not apparent after Coomassie Blue staining, probably due to its smaller predicted size (13.2 kDa), but was clearly identified on Western blots using an anti-T antibody. Both cleavage products were observed

Figure 6. N-cleavage catalyzed by the I_C protein. (A) Schematic illustration of the controllable N-cleavage using the *Ssp* DnaB S1 split-intein. The recombinant precursor protein consists of the 11-aa I_N sequence (along with five native N-extein residues) embedded between a maltose-binding protein sequence (M) and a thioredoxin sequence (T). The IcH protein consists of the 144-aa I_C sequence fused to a His-tag (H, 6 histidines) through a Ser residue, and the Asn residue at the C-terminus of I_C is changed to Ala to prevent splicing. (B) SDS-PAGE analysis of the N-cleavage reaction. Protein bands were visualized by staining with Coomassie blue or by Western blotting using an anti-T antibody against thioredoxin, as indicated. Positions of the precursor protein (MI_NT), the IcH protein, and the cleavage products (M and I_NT) are marked. Lanes 1 and 2 show the partially purified precursor and IcH proteins, respectively. Lanes 3–7 are N-cleavage products after the precursor protein was incubated with the IcH protein alone, with DTT alone, or with the IcH protein plus DTT, as indicated. The IcH protein was used in 8-fold molar excess of the MI_NT precursor protein. (C) Kinetic analysis of the N-cleavage reaction. The plots show time courses of the percentage of the precursor protein that underwent N-cleavage (estimated from Western blots). The precursor protein was incubated with IcH alone (□), with IcH plus 10 mM DTT (◻), or with IcH plus 100 mM DTT (■). The IcH protein was added in ~25-fold molar excess to achieve a pseudo-first order reaction. All experiments were performed in triplicate, and error bars represent standard deviations.

Figure 7. Effect of different protein sequences downstream of I_N on N-cleavage. Purified precursor proteins containing the MI_N portion of the $MI_N T$ protein followed by the indicated protein sequences were incubated with $I_C H$ protein (in 5-fold molar excess) and/or 10 mM DTT for 16 h at room temperature, and reactions were analyzed by SDS-PAGE and Coomassie staining. The positions of precursor proteins, the N-terminal N-cleavage fragment M, and $I_C H$ protein are marked. Bands highlighted with an asterisk correspond to the C-terminal cleavage fragments representing the proteins given above the lanes. Molecular weights of protein standards are given in kDa.

after the precursor protein was treated with the $I_C H$ protein, indicating that the expected N-cleavage had occurred. A higher amount of the N-cleavage was observed when the precursor protein was treated with the $I_C H$ protein in the presence of the reducing agent DTT, although treating the precursor protein with DTT alone did not cause any cleavage. Over 90% of the precursor protein was converted to the cleavage products when an excess amount of the $I_C H$ protein was used together with DTT [Fig. 6(C)].

Rate constants of the above N-cleavage reactions were determined through kinetic analysis [Fig. 6(C)]. A pseudo-first order reaction was achieved by using ~25-fold molar excess of the $I_C H$ protein relative to the precursor protein. The percentage of N-cleavage, measured as the percentage of the precursor protein that had been cleaved, was followed over a period of 24 h at a series of time points. The rate constant (k_{obs}) was calculated to be $(1.9 \pm 2.0) \times 10^{-5} s^{-1}$ when the precursor protein was treated with the $I_C H$ protein alone, $(0.8 \pm 0.1) \times 10^{-4} s^{-1}$ when the precursor protein was treated with the $I_C H$ protein plus 10 mM DTT, and $(2.0 \pm 0.5) \times 10^{-4} s^{-1}$ when the precursor protein was treated with the $I_C H$ protein plus 100 mM DTT.

It was intriguing that the $I_C H$ protein could associate with the I_N sequence in the precursor protein $MI_N T$ to achieve the N-cleavage, considering that the 11-aa I_N sequence was sandwiched between the two globular proteins M and T. We therefore tested whether the N-cleavage reaction succeeded only

because the T (thioredoxin) protein (12 kDa) in the $MI_N T$ precursor was smaller than the $I_C H$ protein (17.5 kDa). We replaced thioredoxin (T) in the $MI_N T$ protein with four protein sequences that are larger than $I_C H$, which included GST (27 kDa), RFP (26.5 kDa), eGFP (28 kDa), and β -galactosidase (β GAL, 116 kDa). When these precursor proteins were purified from *E. coli* and incubated in the presence of $I_C H$ and 10 mM DTT, protein products indicative of N-cleavage were clearly observed on Coomassie-stained SDS gels for all of the four precursor proteins (Fig. 7), although the efficiency of the N-cleavage was relatively low. This relatively low efficiency of N-cleavage was likely due to the suboptimal molar ratio (1:5) between precursor protein and $I_C H$ protein used in this experiment, because other experiments [e.g., Fig. 6(C)] had used a higher molar ratio of 1:25.

Discussion

Mechanistic and structural implications

We have successfully demonstrated that controllable cleavages of recombinant proteins can be achieved using the engineered *Ssp* DnaB S1 split-intein. One part of the split-intein is embedded in the precursor protein, and the complementary part of the split-intein is added *in trans* when needed to activate the desired cleavage reaction. In the C-cleavage design, the 144-aa I_C sequence in the precursor protein could not undergo C-cleavage on its own, which demonstrates for the first time that the 11-aa I_N sequence from the N-terminus of the intein is required for C-cleavage (through Asn cyclization) at the C-terminus of the intein. The I_N sequence is not known to participate directly in catalyzing Asn cyclization; therefore, its effect is more likely structural. In the intein crystal structure, I_N is located together with the C-terminal part (including the Asn residue) of the intein in a catalytic pocket²⁰; therefore, the deletion of I_N might have left a structural void in the catalytic pocket and thus affected cyclization of Asn. The I_C was activated to undergo C-cleavage when the missing I_N was added as a synthetic peptide, indicating that the I_N can associate with the I_C *in trans* to reconstitute an active intein. The I_N peptide lacked an N-extein; therefore, the C-cleavage reaction (Asn cyclization) must have occurred without the first two steps (acyl shift and *trans*-esterification) of the protein splicing mechanism.

In the N-cleavage design, the 11-aa I_N sequence was embedded in the precursor protein, and the separately produced I_C protein must recognize and associate with the short I_N sequence *in trans*, in order for N-cleavage to occur. The reconstituted intein (I_C plus I_N) must have catalyzed an N-S acyl shift at the N-terminus of I_N , and the resulting ester bond was hydrolyzed to complete the N-cleavage reaction. DTT increased the rate constant and the efficiency of the N-cleavage reaction by ~10-fold and more than 4-fold,

Figure 8. Structural modeling of I_C association with I_N for the N-cleavage. (A) Schematic representation of the precursor protein before and after association with the I_C protein. Ribbon structures of I_N (red) and I_C (yellow) are adapted from a crystal structure of the *Ssp* DnaB mini-intein that was split into the I_N and I_C parts, with some of the 12 β -strands numbered according to the original crystal structure.²⁰ The precursor protein consists of the 11-aa I_N (β -strands 1 and 2, red) sandwiched between a maltose-binding protein (MBP) and a thioredoxin (Trx). The 42-kDa MBP and the 12.5-kDa Trx are shown as round balls but not drawn to scale relative to the 17-kDa intein (I_N plus I_C). (B) A hypothetical model for association of I_C with I_N . In the first step, a transient conformational change of the I_C protein loosens its structure into an open-horseshoe shape. In the second step, the I_C protein saddles onto I_N and closes up to form the disk-like structure of an active intein.

respectively, which is likely due to DTT promoting thiolysis of the ester bond, which occurs more rapidly than cleavage by hydrolysis.

The success of our N-cleavage design also for the first time reveals an interesting structural flexibility of the I_C protein, as the I_C protein could functionally assemble with the 11-aa I_N sequence even when the I_N was sandwiched between two large protein domains. The crystal structure of the *Ssp* DnaB mini-intein, from which the S1 split-intein was derived, is shaped like a disk or closed-horseshoe,²⁰ and the 11-aa I_N sequence runs almost perpendicularly through the center of the disk-like structure [Fig. 8(A)]. The 11-aa I_N sequence forms two small β -strands named β_1 and β_2 . The β_2 part of I_N interacts with the β_3 part of the I_C protein to form an antiparallel β -sheet, which may contribute to the association between I_N within the precursor protein and the I_C protein. The β_1 part of I_N is buried deeply inside the intein structure and is enclosed by three β -strands (β_5 , β_6 , and β_{10}) of the I_C protein. It appears spatially impossible for the I_N sequence to simply insert or thread itself through the central cavity of the I_C protein, because in the precursor proteins tested, the I_N sequence was sandwiched between two relatively large globular protein domains (the 42-kDa maltose-binding protein at the N-terminus, and proteins ranging from the 12.5-kDa thioredoxin to the 116-kDa β -galactosidase at the C-terminus). A more likely scenario is that the I_C protein is structurally flexible enough to open up like a clamp so that it could saddle onto the I_N sequence within the

precursor protein, and then close around the I_N sequence to form the active intein [Fig. 8(B)]. It is also possible that the I_C protein preexists in an open-clamp structure and can change to the closed-disk-like structure only upon association with the I_N sequence. The structural flexibility of the intein could be manifested by the two long β -strands (β_5 and β_{10}), which together form the backbone of the disk-like structure of the intein.

Our ability to precisely control the N-cleavage and the C-cleavage reactions permitted easy analysis of the reaction kinetics. Interestingly, the N-cleavage rate constant of this *Ssp* DnaB S1 split-intein is significantly lower than the previously reported N-cleavage rate constants of $(1.0 \pm 0.5) \times 10^{-3} \text{ s}^{-1}$ for the *Ssp* DnaE So split-intein²¹ and $1.9 \times 10^{-3} \text{ s}^{-1}$ for the *Sce* VMA So split-intein.²² This may be due to the very different amino acid sequences of these inteins, although the crystal structures of inteins are generally highly conserved. The different rate constants may be more likely due to the different split sites of these inteins, because the *Ssp* DnaB S1 split-intein is split at the S1 site near the N-terminus of the intein sequence, whereas the *Ssp* DnaE split-intein and the *Sce* VMA split-intein are split at the So site closer to the middle of the intein sequence. Different split sites have been noted to affect the rate constant of *trans*-splicing reactions. The *Ssp* DnaB So split-intein and the *Ssp* DnaB S1 split-intein, which were derived from the same natural intein, showed *trans*-splicing rate constants of $(9.9 \pm 0.8) \times 10^{-4} \text{ s}^{-1}$ and $(4.1 \pm 0.2) \times 10^{-5} \text{ s}^{-1}$,

respectively,^{19,22} which differed by more than an order of magnitude. Unlike the N-cleavage rate constant discussed above, the C-cleavage rate constants of this *Ssp* DnaB S1 split-intein are quite comparable to the previously reported C-cleavage rate constants of $(1.9 \pm 0.9) \times 10^{-4} \text{ s}^{-1}$ for the natural *Ssp* DnaE split-intein and $(1.0\text{--}1.7) \times 10^{-3} \text{ s}^{-1}$ for the synthetic *Sce* VMA split-intein,^{21–23} although these split-inteins have very different amino acid sequences and very different split sites. We found that the C-cleavage rate constant could be increased by ~10-fold by substituting the Cys1 residue to the less nucleophilic Ser residue. Although the residues at the N-terminal splice junction are not known to directly participate in Asn cyclization, it appears that even small changes in amino acid side chains at the intein N-terminus can substantially affect rearrangement of chemical bonds at the C-terminal splice junction.

Potential uses and advantages of the controllable cleavages

Advantages of intein-based protein cleavage methods, compared to others such as protease-based methods, have been noted previously,¹¹ and our methods using intein fragments retain many of these advantages. For example, the N-cleavage method may be used to generate an activated thioester at the C-terminus of a target protein so that the target protein can be joined with another protein or peptide having an N-terminal Cys residue, using the EPL method.^{8–10} Unlike protease-based methods that cleave on the C-terminal side of specific recognition sequences, our intein-based N-cleavage method cleaves on the N-terminal side of the recognition sequence (I_N) and thus allows removal of the affinity purification domain (or tag) placed on the C-terminus of the target protein. The intein-based N- and C-cleavage methods may also be used together on a single target protein to produce precise and tag-free ends at both the N- and the C-termini, or to achieve cyclization of the target protein (ligation of the N- and C-termini) using the EPL approach. We have demonstrated that the C-cleavage method can be used to generate an N-terminal Cys residue on a target protein, which is needed for EPL, although the *Ssp* DnaB intein is naturally followed by a Ser residue. Another advantage of these intein-based methods, unlike some protease-based methods, is that inteins are believed to pose no risk of nonspecific cleavage at unintended locations.

Compared to existing intein-based methods that use contiguous inteins, our methods using the S1 split-intein completely avoid any spontaneous cleavage during expression and purification of the precursor protein. This is a significant advantage, because spontaneous cleavages often result in lower yields of the purified protein. In previous reports using contiguous inteins, unwanted spontaneous cleavages have been observed *in vivo* at various levels and could be as high

as 90%.⁹ In our C-cleavage method, the 11-aa I_N peptide may not be overly expensive and laborious to use, due to the small size of the peptide. For example, cleavage of 100 mg of a 50 kDa-precursor protein may require as little as 20 mg of the I_N peptide, if the peptide is used at ~10-fold molar excess over the precursor protein to drive the cleavage reaction. The extra cost of 20 mg I_N may easily be compensated by an increase in protein yield due to the prevention of spontaneous cleavage observed with other intein-based methods. This cost-effectiveness may be particularly true when producing high value proteins for research or pharmaceutical uses, when using more expensive producing cells (e.g. mammalian tissue culture), or when spontaneous cleavages are not tolerated. Lastly, the small I_N peptide can be stably stored and easily removed from the cleaved proteins through simple dialysis.

The recombinant precursor proteins in this study had either a maltose-binding protein or a chitin-binding domain as the affinity binder for easy purification, but potentially other affinity binders such as the His-tag and the GST-tag may be used. We showed that the C-cleavage can also occur when the precursor protein is bound to chitin beads and incubated with the I_N peptide; therefore, this C-cleavage method may be used in a single-step purification of recombinant proteins, using a process similar to that of the IMPACT method.⁵ In such a process, cell lysate containing the precursor protein is passed through an affinity column, unbound proteins are washed away, the I_N peptide is added to the column to activate the C-cleavage, and the protein of interest is released (cleaved) from the column in a pure form. We have shown the feasibility of this purification strategy by preparing two recombinant proteins (thioredoxin and eGFP) in good yields (6 and 3 mg/L of culture). Our demonstration of peptide-triggered C-cleavage with various different target proteins in solution further shows that the purification approach could be a valuable tool for the preparation of recombinant proteins.

Both N- and C-cleavage reactions presented here reached efficiencies of >90% under optimal conditions, which is comparable to or higher than the cleavage efficiencies of previous methods using contiguous inteins. The N-cleavage method in this study also showed a much higher (~95%) cleavage efficiency compared to the previously reported IMPACT method using contiguous inteins,⁵ probably because the shorter I_N sequence embedded in the precursor protein less likely causes protein misfolding.

Materials and Methods

Plasmid construction

To construct plasmid pM_{IC}T for C-cleavage, the I_C coding sequence of *Ssp* DnaB intein and the thioredoxin coding sequence were PCR-amplified from the

pMST plasmid,¹⁶ and these coding sequences were fused into the pMAL-c2X plasmid (New England Biolabs) between *Xmn* I and *Hind* III sites. Plasmid pMI_CT was constructed by PCR-amplification of the chitin-binding domain coding sequence from the pTWIN1 plasmid (New England Biolabs) and insertion of this sequence at the unique *Age* I site of pMI_CT. Plasmid pMI_NT for N-cleavage was constructed by deleting the I_C coding sequence from the pMST-S1 plasmid¹⁸ through inverse PCR. The genes for β -galactosidase, GST, RFP, and eGFP were cloned into pMI_CT and pMI_NT between *Age* I and *Hind* III sites. Plasmid pTI_CH was constructed by PCR amplification of the I_C coding sequence (with an Asn154Ala mutation and a C-terminal His-tag) from pK2S-M30 (Liu XQ, unpublished) and insertion of this sequence in the pTWIN1 plasmid between engineered *Nde* I and *Hind* III sites. Plasmid pTI_CT' was constructed by fusing the coding sequences of the first 20 residues of maltose-binding protein, the chitin-binding domain, the I_C, and thioredoxin, in this order. The resulting fusion gene was introduced into the pTWIN1 plasmid between engineered *Nde* I and *Hind* III sites. Plasmid pTI_CE was constructed similarly with the sequence of eGFP replacing the thioredoxin sequence and omission of the coding sequence for the first 20 residues of maltose-binding protein.

Protein expression and purification

Precursor proteins containing N-terminal maltose-binding protein were purified using amylose resin (New England Biolabs) affinity chromatography according to manufacturer's instructions. Briefly, *E. coli* DH5 α cells harboring a specified plasmid were grown in 75 mL Luria Broth (LB) to an OD of 0.6, and 0.8 mM IPTG was added to induce protein expression for 3 h at 37°C or over night at room temperature. Cells were harvested by centrifugation, resuspended in amylose column buffer (ACB: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl), and lysed by passing through a French Press (14,000 PSI), all at 4°C. After removing cell debris by centrifugation (10,000 rpm, 25 min), the cell lysate was mixed with 1 mL amylose resin (pre-equilibrated with ACB) for 1 h at 4°C. The resin was then poured into a disposable column (Biorad) and washed with 20 mL ACB. The bound protein was eluted with ACB containing 10 mM maltose and collected in 500- μ L fractions. Purity of the eluted protein was assessed by SDS-PAGE and protein concentration was determined using the Bradford assay (Biorad). The His-tagged I_CH protein was expressed in *E. coli* BL21(DE3) cells (GeneChoice) as described above, except that protein expression was induced with 0.4 mM IPTG at room temperature for 16–18 h. Cells were harvested by centrifugation, resuspended in a binding buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 5 mM imidazole), and lysed using French Press (14,000 PSI). After removing cell debris by centrifuga-

tion, the soluble fraction was filtered (0.45 μ m) and incubated with 1 mL Ni-NTA resin (QIAGEN) for 1 h at 4°C. The resin was poured into a disposable column and washed with 20 mL binding buffer, followed by washing with 15 mL wash buffer (binding buffer plus 60 mM imidazole). Elution was performed with 10 mL strip buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 100 mM EDTA). The eluted protein was dialyzed extensively against optimized splicing buffer (oSB: 20 mM Tris-HCl, pH 8.5, 250 mM NaCl, 1 mM EDTA), and proteins were concentrated by centrifugation using Amicon Ultra Centrifugal Filter Devices (Millipore).

In vitro C-cleavage

Standard reactions for peptide-induced C-cleavage contained ~ 5 μ M MI_CT protein and 200 μ M peptide I_N or I_N2 in cleavage buffer (CB: 100 mM Tris-HCl, pH 7.5, 500 mM NaCl), with or without the addition of reducing agent (50 mM DTT, 0.1 mM TCEP), at 37°C. For kinetic analysis, samples were removed at specific time points, and the reaction was stopped by addition of reducing SDS-PAGE sample buffer. Samples were analyzed by SDS-PAGE on a 12.5% NEXT gel (Mandel Scientific) in combination with a conventional 4% Laemmli stacking gel, followed by staining with Coomassie Blue. Amounts of the precursor protein MI_CT and the cleavage fragment MI_C were estimated from the Coomassie-stained SDS gels through densitometry analysis using ImageJ 1.342. Cleavage efficiencies were defined as the percentage of MI_C over the total (MI_C + MI_CT). Efficiencies were plotted as a function of time, and rate constants (k_{obs}) were determined as described²¹ using KaleidaGraph 4.02.

Production of thioredoxin with N-terminal Cys and purification of eGFP

The TI_CT' and TI_CE precursor proteins were expressed in liquid culture of *E. coli* strain BL21(DE3)pLysS harboring plasmids pTI_CT' and pTI_CE, respectively, which was induced with 0.8 mM IPTG for 18 h at room temperature. Cells were harvested and lysed in 2 mL CB as above. The soluble cell lysate was added to an equal volume of chitin resin (New England Biolabs) on a column. After washing with 18 mL CB, a small sample of the resin was taken, and the remaining resin was soaked with a CB solution containing 100 μ M I_N2 peptide. After incubation for 24 h at 4°C, proteins released from the column were eluted with 2 mL CB.

In vitro N-cleavage

Standard reactions for I_C protein-induced N-cleavage contained ~ 5 μ M MI_NT protein and ~ 100 μ M I_CH protein in oSB, with or without DTT ($c_{\text{final}} = 10$ or 100 mM), at room temperature. Kinetic analysis was performed as described above by measuring the amount of MI_NT precursor and the I_NT fragment, which were visualized by Western blotting using rabbit

anti-T antibody (Sigma) in combination with a secondary mouse anti-rabbit antibody (Sigma) and the Enhanced Chemi-Luminescence detection kit (GE Healthcare).

References

- Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M (1994) Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucleic Acids Res* 22:1125–1127.
- Saleh L, Perler FB (2006) Protein splicing in cis and in trans. *Chem Rec* 6:183–193.
- Xu MQ, Perler FB (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J* 15:5146–5153.
- Paulus H (2000) Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 69:447–496.
- Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* 192:271–281.
- Chong S, Montello GE, Zhang A, Cantor EJ, Liao W, Xu MQ, Benner J (1998) Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res* 26:5109–5115.
- Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB (1999) Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *Biotechniques* 27:110–114, 116, 118–120.
- Evans TC, Jr, Benner J, Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 7:2256–2264.
- Mathys S, Evans TC, Chute IC, Wu H, Chong S, Benner J, Liu XQ, Xu MQ (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* 231(1/2):1–13.
- Evans TC, Jr, Benner J, Xu MQ (1999) The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J Biol Chem* 274:18359–18363.
- Xu MQ, Evans TC, Jr (2001) Intein-mediated ligation and cyclization of expressed proteins. *Methods* 24:257–277.
- Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M (1999) A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol* 17:889–892.
- Cui C, Zhao W, Chen J, Wang J, Li Q (2006) Elimination of in vivo cleavage between target protein and intein in the intein-mediated protein purification systems. *Protein Expr Purif* 50:74–81.
- Wu H, Hu Z, Liu XQ (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci USA* 95:9226–9231.
- Caspi J, Amitai G, Belenkiy O, Pietrokovski S (2003) Distribution of split DnaE inteins in cyanobacteria. *Mol Microbiol* 50:1569–1577.
- Wu H, Xu MQ, Liu XQ (1998) Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim Biophys Acta* 1387(1/2):422–432.
- Lew BM, Mills KV, Paulus H (1999) Characteristics of protein splicing in trans mediated by a semisynthetic split intein. *Biopolymers* 51:355–362.
- Sun W, Yang J, Liu XQ (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J Biol Chem* 279:35281–35286.
- Ludwig C, Pfeiff M, Linne U, Mootz HD (2006) Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing. *Angew Chem Int Ed Engl* 45:5218–5221.
- Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, Lesage G, Rao Z (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J Biol Chem* 278:39133–39142.
- Martin DD, Xu MQ, Evans TC, Jr (2001) Characterization of a naturally occurring trans-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry* 40:1393–1402.
- Brenzel S, Kurpiers T, Mootz HD (2006) Engineering artificially split inteins for applications in protein chemistry: biochemical characterization of the split *Ssp* DnaB intein and comparison to the split *Sce* VMA intein. *Biochemistry* 45:1571–1578.
- Nichols NM, Evans TC, Jr (2004) Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split *Ssp* DnaE intein. *Biochemistry* 43:10265–10276.