Conventional split inteins have been useful for trans-splicing between recombinant proteins, and an artificial S1 split intein is useful for adding synthetic peptide onto the N terminus of recombinant proteins. Here we have engineered a novel S11 split intein for trans-splicing synthetic peptide onto the C terminus of recombinant proteins. The C-intein of the S11 split intein is extremely small (6 amino acids (aa)); thus it can easily be produced together with a synthetic C-extein to be added to the C terminus of target proteins. The S11 intein was derived from the Ssp GyrB intein after deleting the homing endonuclease domain and splitting the remaining intein sequence near the C terminus, producing a 150-aa N-intein (I\(_N\)) and a 6-aa C-intein (I\(_C\)). Its trans-splicing activity was demonstrated first in Escherichia coli cells and then in vitro for trans-splicing between a synthetic peptide and a recombinant protein. The in vitro trans-splicing reaction exhibited a typical rate constant of \((6.9 \pm 2.2) \times 10^{-5}\) s\(^{-1}\) and reached a high efficiency of \(\sim 80\%\). This S11 split intein can be useful for adding any desirable chemical groups to the C terminus of a protein of interest, which may include modified and unnatural amino acids, biotin and fluorescent labels, and even drug molecules.

Inteins are internal protein sequences and catalyze a protein-splicing reaction, which precisely excises the intein sequence and join the flanking sequences (N- and C-exteins) with a peptide bond (1). The reaction mechanism of protein splicing has been well studied (2, 3), and the conserved crystal structure of the protein-splicing domain of inteins consists of \(\sim 12\) \(\beta\)-strands that form a disk-like compact structure with the two splicing junctions located in a central cleft (4–7). More than 400 inteins and intein-like sequences have been found in a wide variety of host proteins and in microorganisms including bacteria, Archaea, and eukaryotes (8, 9). Inteins are typically 350–550 aa\(^3\) in size (8, 9) with the majority containing a homing endonuclease domain. Some inteins are as large as 1650 aa in size and contain tandem repeats (10, 11). Inteins can lose the endonuclease domain, and the resulting functional mini-intein retains only the splicing domain \(\sim 140\) aa in size (12–14).

Split inteins are essentially mini-inteins broken into two pieces and able to reassociate and carry out protein trans-splicing (15). In certain cyanobacteria, a natural split intein is responsible for producing a mature DnaE protein by trans-splicing two separate polypeptides expressed from two separate genes (16, 17). Artificial split inteins have also been engineered by splitting the sequences of contiguous inteins to resemble naturally occurring split inteins (13, 18, 19). Split inteins have many practical uses including the production of recombinant proteins from fragments and the circularization of recombinant proteins (20). However, conventional split inteins are less useful for adding synthetic peptide to proteins because their split sites are relatively close to the middle of the intein sequence and correspond to the insertion site of the homing endonuclease domain. These split sites are referred to as S0 and split the intein sequence into an N-terminal piece of \(\sim 100\) aa and a C-terminal piece of \(\sim 40\) aa (21). These intein pieces are prone to misfolding, and their relatively large sizes make them difficult to be produced through chemical synthesis, particularly after adding desired extein sequences.

Novel split inteins having a split site proximal to the N- or C-terminal would have expanded utilities because they can be used to trans-splice a short synthetic peptide onto recombinant proteins. Recently, an S1 split intein was engineered from the Ssp DnaB intein by splitting the intein sequence at a S1 site proximal to the N-terminal, producing an N-terminal piece (N-intein) of only 11 aa in length and a C-terminal piece (C-intein) of 144 aa in length (21). As the short N-intein can be synthesized chemically with an additional N-extein peptide containing desirable chemical groups, this S1 split intein has been used successfully to trans-splice a fluorophore onto the N terminus of recombinant proteins (22). However, this S1 split intein is not suitable for adding synthetic peptide to the C terminus of proteins because the 144-aa C-intein is too large for chemical synthesis. A previous attempt to produce split inteins having a small C-intein has not been successful (21).

Here we report the successful engineering of a novel S11 split intein for adding synthetic peptide to the C terminus of recombinant proteins. A total of nine natural inteins were first converted to mini-inteins and then split at the S11 split site near the C terminus of the intein sequence, producing an N-intein of \(\sim 150\) aa in length and a C-intein of only 6 aa in length. Among all the inteins tested, an S11 split intein derived from the Ssp GyrB intein was first shown to trans-splice efficiently in Escherichia coli cells. It was further shown to efficiently trans-splice a synthetic peptide with a recombinant protein in vitro. Kinetic analysis of the trans-splicing reaction demonstrated a high efficiency (\(\sim 80\%\)) and a rate constant comparable with those of the
conventional and the S1 split inteins. This novel S11 split intein significantly expands the utility of split inteins, providing the potential to add any number of desirable chemical groups onto the C terminus of target proteins.

**EXPERIMENTAL PROCEDURES**

**Plasmid Construction**—The *Ssp* GyrB mini-intein was constructed by fusing the N- and C-terminal coding sequences of the *Ssp* GyrB intein (8). The N-terminal segment was PCR-amplified from *Synechocystis* sp. PCC6803 genomic DNA using a pair of oligonucleotide primers (5’-CTG GAG GCC GTT TGC TTT TCT GGA GAT ACA TTA GTC GC-3’ and 5’-CAT ATG ACC AGA ATC TTC CGT AGT CGA AAT-3’). The C-terminal segment was obtained similarly, using the primer pair 5’-CAT ATG GAA GCA GTA TTA AAT TAC AAT CAC AG-3’ and 5’-GAC CGG TCT CGC CAG CGC TGT TAT GGA CAA ACA CTC-3’. These segments were digested with appropriate restriction enzymes (XhoI, Ndel, and AgeI) and were placed in the pMST plasmid (21) between the XhoI and Ndel sites. DNA sequencing was used to confirm the sequence of maltose-binding protein as the N-extein and thioredoxin as the C-extein. DNA sequencing was used to confirm the sequence of all plasmids.

To create the *Ssp* GyrB split intein plasmid pMSG-S11, a spacer sequence was inserted into the mini-intein coding sequence at the S11 split site by inverse PCR as described previously (21). The spacer DNA sequentially contains a stop codon, a ribosome binding site, and a start codon and has the following sequence: 5’-TAA TTA ACT TAT AAG GAG GAA AAA CAT ATG-3’. This creates a two-gene operon where the N-protein containing the maltose-binding protein and N-intein is followed by the C-protein containing the C-intein and thioredoxin.

A plasmid (pMSG-S11*inv*) with the N- and C-proteins in the reverse order was modified from pMSG-S11. This was done through circularization of the PCR-amplified two-gene operon from pMSG-S11 followed by linearization at the S11 split site and ligation into a modified pMST-plasmid (lacking extein proteins) at introduced EcoRI and BamHI sites. To produce a plasmid expressing the N-protein only (pMi*N*, H), the coding sequence of a His tag (six histidines) was added to the 3’ end of the iN coding sequence through inverse PCR using the primer pair 5’-CAT CAC CAC CAT CAC CAT TAA TTA ACT TAT AAG GAG GAA AAA CAT ATG-3’ and 5’-GCG TGC CAG AGC AAA ATT GTG-3’, and the C-protein was deleted from the pMSG-S11 plasmid through digestion with Ndel-HindIII followed by filling in and blunt-end ligation.

**Protein Splicing in E. coli cells**—Each expression plasmid was introduced into *E. coli* (DH5α strain). Subsequent protein expression, gel electrophoresis, and Western blotting were carried out as described previously (21). Briefly, cells were grown to late log phase (A600 = 0.5), IPTG was added to a final concentration of 0.8 mM to induce protein expression for 3 h at 37 °C or for ~16 h at room temperature, cells were harvested and lysed in an SDS- and DTT-containing buffer in a boiling water bath, total cellular proteins were analyzed through SDS-PAGE, and protein bands were visualized by staining with Coomassie Brilliant Blue R-250. Western blotting was carried out with either anti-maltose-binding protein monoclonal antibody (New England Biolabs) or anti-thioredoxin monoclonal antibody (Invitrogen), using the enhanced chemiluminescence detection kit (GE Healthcare). The intensity of the protein band was estimated using a gel documentation system (Gel Doc 1000 coupled with Molecular Analyst software, Bio-Rad).

**In Vitro Protein-Peptide trans-Splicing**—The N-precursor protein (Mli*N*, H) was expressed in *E. coli* as above and affinity-purified using amylose resin according to the manufacturer’s instructions (New England Biolabs). The synthetic peptide (L1*F*, F) was purchased from EZBiolab. For in vitro trans-splicing, 20 μM Mli*N*, H protein was incubated with 200 μM L1*F* peptide in a splicing buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA; pH 8.0) in the presence or absence of 0.1 mM TCEP or 1.0 mM DTT at room temperature for a specified length of time. Splicing products were analyzed through SDS-PAGE in the presence of DTT and visualized either by staining with Coomassie Brilliant Blue R-250 or by Western blotting using an anti-maltose-binding protein antibody (see above), an anti-FLAG antibody (Sigma), or an anti-His tag antibody (Roche Applied Science).

**RESULTS**

**Construction of the S11 Split Inteins**—Nine natural inteins were tested to find a functional split site proximal to the C terminus of the intein sequence. They were the *Ssp* GyrB, *Tth* DnaE-1, *Tth* DnaE-2, *Tth* RIR, *CneA* Prp8, *Ter* RIR-1, *Ter* RIR-2, *Ter* RIR-3, and *Ter* RIR-4 inteins. These natural inteins all have contiguous sequences (8), and the last five have previously been shown to be active in cis-splicing (23, 24). Working from the coding sequence of each intein, the homing endonuclease domain sequence (if present) was deleted to produce a mini-intein, based on previously described domain predictions (8, 25, 26). Each mini-intein sequence was then split at the S11 site to produce the split intein. The S11 site was chosen through protein sequence alignments with inteins of known crystal structures such that the S11 site was near the C terminus and between the last two β-strands, β11 and β12. An example of the sequence alignment is shown in Fig. 1, where the *Ssp* GyrB mini-intein is aligned to the *Ssp* DnaB mini-intein, which has a known crystal structure (7).

The S11 split intein constructs were expressed in *E. coli* using the previously described pMST plasmid (21). As illustrated in Fig. 2A, each intein was flanked by a maltose-binding protein (M) as the N-extein and a thioredoxin (T) as the C-extein, expressed from an IPTG-inducible Ptac promoter. For the S11 split intein, the intein coding sequence was split at the S11 site to create a two-gene operon by adding a spacer sequence containing sequentially a stop codon, a ribosome binding site (Shine-Dalgarno sequence), and a start codon. In this operon, the first gene encoded the N-precursor protein consisting of the M sequence (N-extein) followed by the N-terminal intein piece (N-intein or iN), and the second gene encoded the C-precursor protein consisting of the C-terminal intein piece (C-intein or iC) followed by the T sequence as C-extein. Similar gene constructs had been used in previous studies of split inteins (21, 23, 24), which allowed easy expression and identification of the protein products.
Splicing Activities in E. coli—The above plasmids were introduced into E. coli cells to produce the corresponding proteins after IPTG induction, the resulting total cellular proteins were resolved through SDS-PAGE, and relevant protein bands were identified by their predicted sizes and through Western blotting. Proteins were quantified by measuring the intensity of the corresponding signal on Western blots, and the efficiency of splicing was calculated as the ratio of the spliced protein over the total protein (the spliced protein plus the remaining precursor protein). The C-precursor protein was used in estimating the efficiency of trans-splicing because the N-precursor protein existed in excessive amounts due to its higher expression level as the first gene of the two-gene operon, with the exception of the pMSG-S11inv plasmid.

Among all the S11 split intein constructs tested, only the Ssp GyrB S11 split intein showed a trans-splicing activity in E. coli. The spliced protein (MT) was clearly observed after Coomassie Blue staining and confirmed by Western blotting (Fig. 2B), indicating that protein trans-splicing had occurred. A certain amount of the C-precursor protein (ICT) remained, indicating that the trans-splicing reaction did not go to completion. The efficiency (completeness) of the trans-splicing was estimated to be 84% at room temperature but only 16% at 37 °C. The remote possibility of cis-splicing via translational readthrough of the two-gene operon was also investigated. When the gene order in the two-gene operon was reversed in plasmid pMSG-S11inv, making cis-splicing topologically impossible, the spliced protein was still observed (Fig. 2B), thereby ruling out this possibility. Additional tested S11 split inteins, derived from the other eight natural inteins beside the Ssp GyrB intein (see above), did not show a detectable amount of trans-splicing under the tested conditions (data not shown).

FIGURE 1. S11 split site in the Ssp GyrB mini-intein sequence. The amino acid sequence of the Ssp GyrB mini-intein is aligned with that of the Ssp DnaB mini-intein. Identical and similar amino acid residues are marked by : and *, respectively. Gaps (–) are introduced for optimal alignment. These mini-inteins were derived from their respective natural inteins by deleting the predicted homing endonuclease domain sequences, with the position and the number of deleted amino acids indicated by an open arrowhead and in parentheses, respectively. The Ssp DnaB mini-intein has a known crystal structure, and the 12 β-strands (β1 to β12) are underlined. The position of the S11 split site is marked with a solid arrowhead.

FIGURE 2. S11 split intein trans-splicing in E. coli cells. A, schematic illustration of the trans-splicing reaction. The recombinant N-precursor protein consists of a maltose-binding protein sequence (M) followed by the 150-aa N-intein sequence (IN). The recombinant C-precursor protein consists of the 6-aa C-intein sequence (IC) followed by a thioredoxin sequence (T). Coding sequences of these precursor proteins were placed in an expression plasmid as a two-gene operon behind the IPTG-inducible Ptac promoter. In plasmid pMSG-S11, the N-precursor gene precedes the C-precursor gene. In plasmid pMSG-S11inv, the C-precursor gene precedes the N-precursor gene. B, SDS-PAGE analysis of the trans-splicing reaction. Protein bands were visualized by Coomassie blue staining or by Western blotting using an anti-T antibody against the thioredoxin sequence, as indicated. Positions are marked for the N-precursor protein (M), the C-precursor protein (LcT), and the spliced protein (MT). Lanes 1–3 are total proteins of E. coli cells before the IPTG induction of precursor protein expression, after IPTG induction for 3 h at 37 °C, and after IPTG induction overnight at room temperature, respectively. The E. coli cells harbored either the pMSG-S11 plasmid or the pMSG-S11inv plasmid as indicated.
antibodies, the N-terminal intein piece INH was recognized spliced protein MF was recognized by both anti-M and anti-F N-precursor protein with a synthetic peptide (Fig. 3). The N-precursor protein MINH consisted of a maltose-binding protein (M), the N-intein (IN), and a His tag sequence (H). The synthetic peptide (ICF) is 30 aa in length with the following sequence: GVFVHNSADYKDDDDKSGCLAGDTLITLAS. It contains the IC sequence (GVFVH) followed by a C-extein sequence (F) that begins with a Ser residue and contains the FLAG epitope. B, SDS-PAGE analysis of the trans-splicing reaction. Protein bands were visualized by Coomassie Blue staining, Western blot using an anti-M antibody against the maltose-binding protein sequence, Western blot using an anti-H antibody against the His tag sequence, or Western blot using an anti-F antibody against the FLAG epitope, as indicated. Positions are marked for the N-precursor protein (MINH), the spliced protein (MF), and the released N-intein (INH). The N-precursor protein was incubated for 16 h at room temperature alone, with the IC-F peptide, with the IC-F peptide plus 0.1 mM TCEP, or with the IC-F peptide plus 1 mM DTT, as indicated.

**Protein-Peptide trans-Splicing in Vitro**—The Ssp GyrB S11 split intein was tested in vitro to trans-splice a recombinant N-precursor protein with a synthetic peptide (Fig. 3A). The N-precursor protein MINH consisted of a maltose-binding protein (M), the N-intein (IN), and a His tag sequence (H). The synthetic peptide (IC-F) is 30 aa in length with the following sequence: GVFVHNSADYKDDDDKSGCLAGDTLITLAS. It contains the IC sequence (GVFVH) followed by a C-extein, and the C-extein sequence starts with a serine residue required for splicing and contains the FLAG epitope (DYKDDDDK) for easy detection. The trans-splicing occurs, the FLAG-containing C-extein (F) will be transferred onto the C terminus of the maltose-binding protein (M), producing the spliced protein (MF). Simultaneously, the N-intein piece (INH) will be released from the precursor protein, and the C-intein (IC) will be released from the synthetic IC-F peptide.

The predicted products of the above protein-peptide trans-splicing were observed after Coomassie Blue staining and identified through Western blotting (Fig. 3B). As expected, the spliced protein MF was recognized by both anti-M and anti-F antibodies, the N-terminal intein piece INH was recognized only by the anti-H antibody, and the N-precursor protein MINH was recognized by both anti-M and anti-H antibodies. This trans-splicing reaction was enhanced by the addition of reducing agents (TCEP and DTT), as indicated by the increased amounts of the spliced protein (Fig. 3B). Over 80% of the N-precursor protein was converted into the spliced protein in the presence of 1 mM DTT after 11 h of incubation with the synthetic peptide (Fig. 4A). A minor protein band corresponding to the M part of the N-precursor protein (MINH) was also observed, suggesting that a small amount of the N-precursor protein underwent N-cleavage (i.e. peptide bond breakage at N terminus of the IN sequence) instead of trans-splicing.

The rate constant of the above trans-splicing was also determined for comparison with conventional split inteins. The purified N-precursor protein was incubated with a 10-fold molar excess of the IC-F peptide to achieve a pseudo-first order reaction regarding the N-precursor protein. The reaction products were resolved through SDS-PAGE, visualized by Coomassie Blue staining, and quantified through laser scanning and gel documentation. The amount of the spliced product was calculated as a percentage of the starting amount of N-precursor protein. The percentage of spliced product formation was plotted against time from a series of measurements over a 31-h period (Fig. 4B). The plot was fitted to the pseudo-first order reaction equation of $p = P_0(1-e^{-kt})$ (28), and the rate constant ($k_{\text{obs}}$) was determined to be $(6.9 \pm 2.2) \times 10^{-5}$ s$^{-1}$.

**DISCUSSION**

We have succeeded for the first time in constructing a novel S11 split intein capable of protein trans-splicing after testing nine different natural inteins by splitting the intein sequences at the S11 site near the C terminus. For a trans-splicing reaction to occur, the two precursor proteins must be able to recognize and associate with one another. The exteins used in this study, maltose-binding protein and thioredoxin, do not interact structurally and cannot play a role in bringing together the two precursor proteins. Therefore the intein fragments (IN and IC) must be solely responsible for the recognition and association between the two precursor proteins in vivo or between the N-precursor protein and the short peptide in vitro. Interestingly, the functional association between the IN and IC appears to be a rare property of the S11 split intein derived from the Ssp GyrB intein because similarly constructed S11 split inteins derived from nine other natural inteins did not show a trans-splicing activity in this study and in a previous report (21).

It is striking that the 6-aa IC could associate with the 150-aa IN to reconstitute an active intein. The structural basis of this IN-IC association may be considered in terms of the highly conserved crystal structures of inteins in general, although no crystal structure is available for the Ssp GyrB intein. The highly conserved mini-intein structure typically consists of ~12
FIGURE 4. Kinetic analysis of the protein-peptide trans-splicing. A, the trans-splicing reaction analyzed through SDS-PAGE and Coomassie Blue staining. The purified N-precursor protein (M_in_H) was incubated with the synthetic ICF peptide in presence of 1 mM DTT at room temperature for the indicated lengths of time, the reaction products were resolved through SDS-PAGE, and the protein bands were visualized by Coomassie blue staining. Positions of the N-precursor protein (M_in_H) and the spliced protein (MF) are marked. M marks a protein band corresponding to a cleavage product of the N-precursor protein.

β-strands (4–7), and the 6-aa I_C sequence corresponds to the last β-strand β12 in the Ssp DnaB mini-intein (Fig. 1). In the disk-like structure of the mini-intein, β12 is located inside the centrally positioned catalytic pocket of the intein, and it interacts directly with β11 to form a short anti-parallel β-sheet. This may explain the intermolecular association between I_C and I_C of the S11 split intein because the I_C contains β12, whereas the I_N contains β11 and other parts of the catalytic pocket.

The in vitro trans-splicing conditions allowed for an analysis of the reaction kinetics of the Ssp GyrB S11 split intein. The observed rate constant of $(6.9 \pm 2.2) \times 10^{-5}$ s$^{-1}$ is close or comparable with previously reported rate constants of the conventional and the S1 split inteins. The Ssp DnaB S1 split intein, whose split site is near the N-terminal of intein (21), exhibited a rate constant of $(4.1 \pm 0.2) \times 10^{-5}$ s$^{-1}$ (22). Conventional split inteins, whose split sites are near the middle of the intein and correspond to the insertion site of the homing endonuclease domain, exhibited rate constants ranging from $\sim 10^{-4}$ s$^{-1}$ for the synthetic Ssp DnaB split intein (27) to $\sim 10^{-5}$ s$^{-1}$ for the naturally occurring Ssp DnaE split intein (28).

In addition to trans-splicing, a small amount of N-cleavage (at the N terminus of I_N) was observed for the Ssp GyrB S11 split intein when DTT was present. This is not surprising because strong nucleophilic compounds such as DTT are known to resolve ester bonds formed during the first two steps (N–S/
12. Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W. R., Jr., and Perler, F. B. (1997) *J. Bacteriol.* **179**, 6378–6382
13. Wu, H., Xu, M. Q., and Liu, X. Q. (1998) *Biochim. Biophys. Acta* **1387**, 422–432
14. Derbyshire, V., Wood, D. W., Wu, W., Dansereau, J. T., Dalgaard, J. Z., and Belfort, M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11466–11471
15. Liu, X. Q. (2000) *Annu. Rev. Genet.* **34**, 61–76
16. Wu, H., Hu, Z., and Liu, X. Q. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9226–9231
17. Caspi, J., Amitai, G., Belenkiy, O., and Pietrokovski, S. (2003) *Mol. Microbiol.* **50**, 1569–1577
18. Mills, K. V., Lew, B. M., Jiang, S., and Paulus, H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3543–3548
19. Southworth, M. W., Adam, E., Panne, D., Byer, R., Kautz, R., and Perler, F. B. (1998) *EMBO J.* **17**, 918–926
20. Evans, T. C., Jr., Martin, D., Kolly, R., Panne, D., Sun, I., Ghosh, I., Chen, L., Benner, J., Liu, X. Q., and Xu, M. Q. (2000) *J. Biol. Chem.* **275**, 9091–9094
21. Sun, W., Yang, J., and Liu, X. Q. (2004) *J. Biol. Chem.* **279**, 35281–35286
22. Ludwig, C., Pfeiff, M., Linne, U., and Mootz, H. D. (2006) *Angew Chem. Int. Ed. Engl.* **45**, 5218–5221
23. Liu, X. Q., Yang, J., and Meng, Q. (2003) *J. Biol. Chem.* **278**, 46826–46831
24. Liu, X. Q., and Yang, J. (2004) *FEBS Lett.* **572**, 46–50
25. Pietrokovski, S. (1998) *Protein Sci.* **7**, 64–71
26. Dalgaard, J. Z., Moser, M. J., Hughey, R., and Mian, I. S. (1997) *J. Comput. Biol.* **4**, 193–214
27. Brenzel, S., Kurpiers, T., and Mootz, H. D. (2006) *Biochemistry* **45**, 1571–1578
28. Martin, D. D., Xu, M. Q., and Evans, T. C., Jr. (2001) *Biochemistry* **40**, 1393–1402