An Unprecedented Combination of Serine and Cysteine Nucleophiles in a Split Intein with an Atypical Split Site*

Anne-Lena Bachmann and Henning D. Mootz

Background: Intein-mediated protein splicing involves rearrangement of the polypeptide backbone through ester intermediates and excises the intein from the precursor.

Results: An unusual split intein was identified that converts an oxoester into a thioester.

Conclusion: Inteins can catalyze two consecutive formally uphill reactions.

Significance: The first example of this special case in protein splicing underlines the mechanistic diversity of inteins.

Protein splicing mediated by inteins is a self-processive reaction leading to the excision of the internal intein domain from a precursor protein and the concomitant ligation of the flanking sequences, the extein-N and extein-C parts, thereby reconstituting the host protein. Most inteins employ a splicing pathway in which the upstream scissile peptide bond is consecutively rearranged into two thioester or oxoester intermediates before intein excision and rearrangement into the new peptide bond occurs. The catalytically critical amino acids involved at the two splice junctions are cysteine, serine, or threonine. Notably, the only potential combination not observed so far in any of the known or engineered inteins corresponds to the transesterification from an oxoester to a thioester, which suggested that this formal uphill reaction with regard to the thermodynamic stability might be incompatible with intein-mediated catalysis. We show that corresponding mutations also led to inactive gp41-1 and AceL-TerL inteins. We report the novel GOS-TerL split intein identified from metagenomic databases as the first intein harboring the combination of Ser1 and Cys + 1 residues. Mutational analysis showed that its efficient splicing reaction indeed follows the shift from oxoester to thioester and thus represents a rare diversion from the canonical pathway. Furthermore, the GOS-TerL intein has an atypical split site close to the N terminus. The IntN fragment could be shortened from 37 to 28 amino acids and exchanged with the 25-amino acid IntN fragment from the AceL-TerL intein, indicating a high degree of promiscuity of the IntC fragment of the GOS-TerL intein.

Inteins are internal protein elements that excise themselves out of a precursor protein in a self-processive reaction called protein splicing (Fig. 1a; for reviews see Ref. 1–4). The flanking N- and C-terminal extein sequences are concomitantly linked with a peptide bond (5, 6). Protein splicing is a remarkable multistep pathway involving a series of bond rearrangements and nucleophilic displacements (Fig. 1b) (7–10). Although the overall protein structure and mechanistic logic is very similar for all inteins, different classes can be defined according to major differences in the protein splicing pathway. The class 1 inteins constitute the largest group of inteins and employ the canonical protein splicing pathway (Fig. 1) (9, 11). Here, the scissile peptide bond upstream of the intein between the –1 extein residue and the first amino acid of the intein (residue 1), which can be a cysteine or a serine, is first rearranged into the thioester or oxoester by an N-S or N-O acyl shift, respectively. This linear ester intermediate with higher energy is then attacked by the nucleophilic side chain of the first residue downstream of the intein, the +1 residue, which can be a cysteine, serine, or threonine. Following this transesterification to the branched intermediate, asparagine cyclization at the ultimate position of the intein leads to formation of a C-terminal succinimide ring on the intein and to the cleavage of the downstream scissile peptide bond. The liberated N- and C-terminal extein segments are linked at this point via the side chain ester of the +1 residue, however, rearrange quickly into the stable peptide bond by intramolecular attack of the α-amino group. Class 2 inteins divert from this splicing pathway in that they omit the first step and catalyze instead a direct attack of the +1 nucleophilic side chain on the upstream peptide bond (11, 12). Class 3 inteins also do not show peptide bond rearrangement at the upstream splice junction, but have an additional internal cysteine side chain that forms a thioester with the N-extein fragment through direct attack on the upstream peptide bond, before transesterification to the +1 residue and the further canonical order of events occur (11).

Inteins can be regarded as single turnover catalysts. To catalyze protein splicing, several residues have important or essential functional roles. Most of these residues can be recognized as highly or reasonably well conserved in short signature sequences, designated block A, B, F, and G or motifs N1, N2, N3, C2, C1 (alternative designation) (13–15). We will refer to these as block A/N1, block B/N3 etc. In addition to the different classes of inteins, also more subtle differences regarding the contribution of individual residues to catalysis have been noted. A conserved residue essential in one particular intein may be only important for catalysis in another intein.

It is not understood why some class 1 inteins use cysteine and others serine at the catalytic position 1. Likewise, the reason for using cysteine, serine, or threonine at the catalytic position +1

* This work was supported in part by Deutsche Forschungsgemeinschaft Grants MO 1073-3,2 and SPP1623. The authors declare that they have no conflicts of interest with the contents of this article.

1 To whom correspondence may be addressed. Fax: 49-251-8333007; E-mail: Henning.Mootz@uni-muenster.de.
Intein with Unprecedented Nucleophile Combination

remains enigmatic. Interestingly, however, clear preferences for the different combinations of these key residues can be observed. Liu and co-workers (17) have analyzed the intein database (InBase) (16) and found a bias in the possible combinations of the 1 and +1 positions. The Cys1/Cys+1 combination was the most common one (36%), followed by Cys1/Ser+1 and Cys1/Thr+1 (29 and 12%, respectively). Much rarer were combinations Ser1/Ser+1 and Ser1/Thr+1 employing only oxoester chemistry (both 5%). Strikingly, the combination Ser1 and Cys+1 has not been reported so far in all naturally occurring inteins and it is unclear why this is the case. The findings from naturally occurring inteins were also challenged in several studies using site-directed mutagenesis to artificially change the combinations of the 1 and +1 residues. Importantly, in no case was an efficiently splicing intein with a Ser1/Cys+1 combination reported. For example, a C1S mutation of the highly active Npu DnaE intein (native Cys1/Cys+1 combination) inhibited protein splicing (18). Also the Sce VMA intein (Cys/Cys+1 combination) showed no splicing activity with a C1S mutation (19). In one report the same intein with a combination of C1S and H362N mutation in the block B/N3 motif was described to yield an active intein, however, no data to support this notion was shown (20, 21). Because only short exteins of a few amino acids were used in this study, a gel-based readout could not distinguish between splicing and cleavage reactions. Furthermore, a systematic study with six inteins showed that none of these accepted the Ser1/Cys+1 combination (17). Together, these observations raise the question if a Ser/Cys+1 combination is potentially incompatible with an efficiently splicing intein. The simple chemical explanation would be that a thioester is more energy-rich than an oxoester (22).

In split inteins the intein domain is separated on two polypeptides. For protein trans-splicing to occur, the N-intein (IntN)2 and C-intein (IntC) fragments have to associate and reconstitute into the active intein (Fig. 1). Various split sites within the conserved intein-fold have been created artificially and led to active and biotechnologically useful inteins (23). Only a small fraction of naturally occurring intein is split and until very recently all known inteins exhibited the split point at the same position ~35–40 residues upstream of the C-terminal end of the intein sequence (total size about 130–150 aa) (24). This split point corresponds to the typical insertion site of the homing endonuclease domain found in maxi-inteins. Very recently, the first atypically split intein was discovered in metagenomic databases and biochemically characterized. The Acel-TerL intein consists of a short IntN fragment of 25 aa and a C-terminal fragment of 104 aa (25). Metagenomic databases proved already earlier to be a rich source for new split inteins with interesting properties (26, 27). For example, the gp41-1 intein is the fastest splicing split intein described so far (27).

In this work, we report on the identification and biochemical characterization of a novel naturally split intein, the GOS-TerL intein, identified from metagenomic sources. This intein employs the so far unknown combination of Ser1 and Cys+1 as catalytic residues. Furthermore, it is the second naturally split intein with an atypical split site close to the N terminus. Using site-directed mutagenesis, we establish the role of key residues in the protein splicing pathway. We show that the GOS-TerL intein belongs to the class 1 inteins, however, its diverting splicing pathway involves the unusual shift from an oxoester to a thioester intermediate.

**Experimental Procedures**

General—Unless otherwise specified, standard protocols were used. Oligonucleotides were purchased from Biolegio (Nijmegen, Netherland). All plasmids were confirmed by DNA sequencing (GATC, Konstanz, Germany, and Seqlab, Göttingen, Germany). The plasmids with synthetic DNA were ordered from MrGene (Konstanz, Germany). All reactions were performed at least in triplicates. Reagents were purchased from 5-Prime (Hamburg, Germany), Acros Organics (Niederau, Germany), Agilent (Böblingen, Germany), AppliChem (Darmstadt, Germany), Fluka (Taufkirchen, Germany), GE

2 The abbreviations used are: IntN, N-intein; IntC, C-intein; aa, amino acid(s); MBP, maltose-binding protein; Trx, thioredoxin; Fl, (5,6)-carboxyfluorescein.
Intein with Unprecedented Nucleophile Combination

TABLE 1
Full amino acid sequences of GOS-TerL proteins used in this study

<table>
<thead>
<tr>
<th>Protein</th>
<th>Plasmid name</th>
<th>Encoded amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP-Int32-His6</td>
<td>pAB44</td>
<td>(MBP)-NSSSSSSSSSSSLLGEGRSEFLVPGRGKVEFISSQSEYINIEVNGKVEYIK</td>
</tr>
<tr>
<td>MBP-Int51 (SIC)-His6</td>
<td>pAB49</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>MBP-Int51 (SIC)-His6</td>
<td>pAB47</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-Tx-His</td>
<td>pAB79</td>
<td>See indicated deletion</td>
</tr>
<tr>
<td>Int51-(S87A)-Tx-His</td>
<td>pAB62</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(S87T)-Tx-His</td>
<td>pAB117</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(H90A)-Tx-His</td>
<td>pAB35</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(C86S)-Tx-His</td>
<td>pAB70</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(C117A)-Tx-His</td>
<td>pAB77</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(C117A)-Tx-His</td>
<td>pAB82</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(C117A)-Tx-His</td>
<td>pAB100</td>
<td>See indicated mutations</td>
</tr>
<tr>
<td>Int51-(D112A)-Tx-His</td>
<td>pAB57</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(L152A)-Tx-His</td>
<td>pAB15</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(C+1A)-Tx-His</td>
<td>pAB38</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(C+1A)-Tx-His</td>
<td>pAB12</td>
<td>See indicated mutation</td>
</tr>
<tr>
<td>Int51-(L152A, C+1A)-Tx-His</td>
<td>pAB32</td>
<td>See indicated mutations</td>
</tr>
<tr>
<td>Int51-(H90A, N152A, C+1A)-Tx-His</td>
<td>pAB69</td>
<td>See indicated mutations</td>
</tr>
</tbody>
</table>

a Int51-encoding plasmid were derived from pMAL-c2X (induction with IPTG), whereas Int51-encoding plasmids are based on pBAD (induction with arabinose).

b The Int51 and Int51 fragment are in italics and the mutated and deleted amino acids in the respective constructs are underlined and in bold in the primary sequences encoded by pAB44 and pAB51. MBP, maltose-binding protein.

Healthcare (Munich, Germany), IRIS Biotech GmbH (Marktredwitz, Germany), Novabiochem (Bad Soden, Germany), Roth (Karlsruhe, Germany), or Sartorius (Göttingen, Germany). Restriction enzymes and markers were obtained from Fermentas (St. Leon-Rot, Germany) and New England Biolabs (Ipswich, MA). Murine-derived anti-His antibody was purchased from Roche (Mannheim, Germany; catalogue number 11604905318001) and used in a 1:5,000 dilution; anti-Trx antibody (Deisenhofen, Germany; catalogue number T0803) and used in a 1:20,000 dilution.

GOS-TerL Intein Construction of Expression Plasmids—A piece of synthetic DNA was used to encode the GOS-TerL intein with five native extein residues at each side and containing an optimized codon usage for high level expression in *Escherichia coli* (see native gene sequence in NCBI GenBank™ entry AACY020533749.1). The KVEFE-Int32-His6 encoding fragment was excised from this vector via BamHI and HindIII and ligated into pMAL-c2X backbone to generate MBP-Int32-His6. In a similar manner, the Int51-CEFLG encoding fragment was inserted in pVS01 (28) using the NdeI and KpnI restriction sites to give Int51-Tx-His6. Site-directed mutagenesis was performed according to the QuikChange protocol (Stratagene).

Expression and Purification of Proteins—*E. coli* BL21(DE3) cells were transformed with each of the expression plasmids (Table 1 for sequence information of the encoded proteins). The individual expression strains were grown at 37 °C in LB media until A600 = 0.5–0.7 was reached. The temperature was then reduced to 28 °C and 0.4 mM isopropyl 1-thio-β-D-galactopyranoside or 0.2% arabinose (Table 1) was added to induce expression. After 4 h cells were spun down and resuspended in nickel-nitrilotriacetic acid (50 mM Tris/HCl, 300 mM NaCl, pH 8) buffer with 8 mM urea. Denaturing purification conditions led to higher yields for Int51 proteins and higher splicing efficiency of Int51 full-length proteins. Cells were disrupted using a Potter-Elvehjem homogenizer. Proteins were purified as previously described using nickel-nitrilotriacetic acid chromatography (28).

**Protein Splicing Assay**—Protein trans-splicing with purified proteins or peptides was performed by mixing complementary intein fragments in splicing buffer (50 mM Tris, 300 mM NaCl, 1 mM EDTA, and 2 mM DTT, pH 7.0). Reaction partners were at a concentration of 7 to 45 μM with an excess of Int51 if not stated otherwise. Proteins were preincubated at specified temperatures before the reaction was started by mixing the components. At various time points aliquots were removed and the reaction quenched by adding 4× SDS-PAGE loading buffer (containing 8% SDS and 20% β-mercaptoethanol) except for oxoester cleavage assays (see below). Samples were boiled (96 °C, 10 min) and analyzed by SDS-PAGE gels stained with Coomassie Brilliant Blue. Relative intensities of protein bands were densitometrically determined using the program ImageJ.

**Oxoester Cleavage Assay**—Cleavage was performed with 10 μM Int51 protein in a modified splice buffer (50 mM NaH2PO4, 300 mM NaCl, 1 mM EDTA, pH 6) at 30 °C, with Int51 protein (14.6 μM) and DTT, hydroxylamine, or tris(2-carboxyethyl)-phosphine as specified. The reaction was stopped with 2× SDS sample buffer without DTT and β-mercaptoethanol and boiled for 10 min at 96 °C.

**Peptide Synthesis and Purification**—Solid phase peptide synthesis and purification was performed as reported previously (29, 30) using a Liberty peptide synthesizer (CEM, Kamp-Lintfort, Germany) on a Tentagel-Wang resin according to standard protocols based on the Fmoc (N-(9-fluorenylmethoxycarbonyl) protecting group strategy and with 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethylyuronium hexafluorophosphate as the activation reagent in the presence of oximapure. All peptides were purified by preparative HPLC (Knauer Smartline-Series, Germany) under acidic conditions (buffer A: 0.05% TFA in H2O; buffer B: 100% acetonitrile + 0.05% TFA). The starting conditions were 95% buffer A and 5% buffer B. The gradient was run to 20% buffer A and 80% buffer B over 45 min with a VariTide RPC column (Varian, Germany). Eluted peptides were confirmed by MALDI-TOF MS analysis (pep-C1: FI-KKEFE-CVYDGM- VETEDGKIKIEDLYKRLA [M + H]+ calc = 3910.3 Da, [M +
Intein with Unprecedented Nucleophile Combination

H$^+_{\text{obs}}$ = 3912.4 Da; pep-C1S: Fl-KKEFE-SVYGDTMVET-EDGKIKIEDLYKRLA [M + H]$^+$$_{\text{calc}}$ = 3894.9 Da, [M + H]$^+$$_{\text{obs}}$ = 3896.4 Da; Fl is 5-(6)-carboxyfluorescein.

Mass Spectrometry—LC-MS/MS analysis of a peptide extract from an in-gel tryptic digest of the protein treated with iodoacetamide to alkylate cysteine residues was performed on an Ultimate 3000 system (Thermo Fisher Scientific Inc., MA, USA) coupled to a maxis II qTOF (Bruker Corp., MA) via an Apollo II ESI source. A Poroshell 120 EC-C18 4.6 × 50-mm 2.7-μm column (Agilent Technologies Inc., CA) was employed for the separation with a 30-min gradient from 5 to 95% acetonitrile in water in the presence of 0.1% formic acid at a flow rate of 0.6 ml/min. Data manipulation and analysis were carried out with Bruker Compass DataAnalysis 4.3 and Bruker Daltonics BioTools 3.2 (both Bruker Daltonik GmbH, Germany).

Results and Discussion

Identification of the GOS-TerL Intein and Sequence Comparison with the AceL-TerL Intein—We identified a new intein with the unique combination of a Ser1 residue at the upstream splice junction and a Cys+1 residue at the downstream splice junction. By using the amino acid sequence of the AceL-TerL intein (25) as search query this intein was found in the metagenomic database at NCBI. The two fragments were identified manually and the putative intein boundaries were defined by the low homology to other inteins. The sequence data were collected by a global ocean sampling project at the Punta Cormorant hypersaline lagoon project in Galapagos Floreana Island (31). As predicted from the DNA sequence, the new intein consists of 152 aa and is naturally split AceL-TerL intein (129 aa) (25) (Fig. 2). The IntN and IntC sequences show 30 and 33% identity, respectively. The most significant differences appear to be localized at the residue at position 1 and the A/N1 motif. Both inteins are integrated at the same point in very homologous phage terminase large subunits (68% identity). The inteins have identical flanking residues at their immediate N’ (EFE) and C’ (CEFLG) ends. Despite these similarities the origins of the metagenomic samples are very different. The AceL-TerL intein was discovered from environmental samples in Ace Lake in Antarctica, where the average water temperature was 1–2 °C, whereas the DNA encoding the GOS-TerL intein was sampled from a hypersaline lagoon at the Galapagos Islands in Ecuador, where the temperature was 37 °C (31). Both habitats are hypersaline salt water bodies.

To rationalize the split site of the GOS-TerL intein within the conserved HINT-fold of inteins, we predicted the structure with an intensive modeling at the phyre2 server (33). The highest structural similarity was assigned to the Pab PolII intein (Protein Data Bank number 2LCJ) (34) and the Pho RadA intein (Protein Data Bank number 4E2U) (35). As shown in Fig. 3, a split site after 37 aa would correspond to the loop structure on the surface of the intein-fold. This position corresponds to position 24/25 following α-helix 1 reported for the AceL-TerL inteins (25) (see below). The GOS-TerL intein thus represents the second example for this natural, atypical split site.

In Vivo Splicing in Cis and in Trans of the GOS-TerL Intein—To functionally characterize the GOS-TerL intein we ordered synthetic DNA deduced from the amino acid sequence with codon optimization for E. coli and cloned it into expression vectors. A gene fragment encoding the artificially fused intein fragments linked by a Gly-Ser dipeptide linker, and with five native N and C terminally flanking residues, KVEFE and CEFGL, respectively, was inserted between maltose-binding protein (MBP) and thioredoxin (Trx) as model exteins (Fig. 4a) (36). Following expression in E. coli, we could identify in total cell extracts large amounts of ligated exteins MBP-Trx and the liberated intein (Fig. 4c, left panels), as well as minor amounts of unspliced precursor. Western blotting using an anti-Trx antibody confirmed this interpretation. Furthermore, the antibody weakly stained a protein that from its molecular weight would...
Individual proteins with a 3-fold excess of MBP-IntN-His6 at 8M urea by dialysis, splice reactions were started by mixing the intC fragment with the intN fragment under native conditions. Following removal of the denaturant, the reaction mixture was analyzed by Western blot analysis. The reaction mixture was found to result in trans-splicing activity, although for the cis-splicing construct a non-native codon usage optimized for E. coli was used (data not shown). To test expression of the IntC-Trx protein in the context of the natural crystal structure of the Pab PolII intein (Protein Data Bank 2LCJ) (34) was performed. The sequences of the IntN and IntC fragments were fused directly without any linker amino acids. IntN is shown in red and IntC in gray. The extruding exteins are shown in green and the N and C termini are indicated, when visible in the structure. The Ser1 and Cys+1 residue are highlighted in stick representation. The split site is marked with an arrow in the predicted structure, A, top view. B, side view.

We wondered whether this finding resulted from the split nature of the intein and reflected an internal start of translation, although for the cis-splicing construct a non-native codon usage optimized for E. coli was used (data not shown). To test expression of the IntC-Trx protein in the context of the natural crystal structure of the Pab PolII intein (Protein Data Bank 2LCJ) (34) was performed. The sequences of the IntN and IntC fragments were fused directly without any linker amino acids. IntN is shown in red and IntC in gray. The extruding exteins are shown in green and the N and C termini are indicated, when visible in the structure. The Ser1 and Cys+1 residue are highlighted in stick representation. The split site is marked with an arrow in the predicted structure, A, top view. B, side view.

We first investigated the upstream splice junction (Fig. 6a). The S1A mutation completely blocked splicing activity and allowed formation of only minor amounts of C-terminal cleavage resulting from uncoupled asparagine cyclization. Substitution of this residue with cysteine (S1C mutant) supported full splicing activity, although at a slightly slower rate than with the wild-type sequence (Fig. 6b). To more directly test if a linear (oxo)ester was formed as an intermediate we blocked all downstream reactions by introducing the mutations N152A and C+1A at the C-terminal splice junction. The expected (oxo)ester formed at the upstream scissile bond was then probed for nucleophlic cleavage using DTT and hydroxylamine, both of which are known to cleave thioesters efficiently and oxoesters only poorly (9). Reducing conditions were found to be required to resolve efficient formation of disulfide-linked homodimers in case of the MBP-IntN(SIC)-His6 construct (Cys1 is the only cysteine residue in this protein). As shown in Fig. 6c, a fraction of the wild-type MBP-IntN-His6 precursor protein (with Ser1) was cleaved to give free MBP, both in the absence and presence of nucleophiles including tris(2-carboxyethyl)phosphine as a non-nucleophilic control for reducing conditions. In contrast, the S1A mutant rendered the protein completely resistant against cleavage at the scissile bond (data not shown). Importantly, cleavage was strictly dependent on the presence of complementary construct IntC(N152A,C+1A)-Trx-His6 underlining that intein reconstitution is required (data not shown). In the case of the S1C construct, cleavage with both nucleophiles DTT and hydroxylamine was clearly more efficient (Fig. 6c), whereas cleavage with tris(2-carboxyethyl)phosphine was comparable with the Ser1 construct. We also prepared the construct IntC(H90A,N152A,C+1A)-Trx-His6 in which the highly conserved histidine of block B was mutated. This histidine was shown in other inteins to be important or essential for (thio)ester formation at the N-terminal scissile bond (30, 37–39). No cleavage could be observed in combination with the wild-type IntN construct (data not shown). Together, these results are in agreement with the typical rearrangement of the peptide bond around Ser1 or Cys1, respectively. Both the oxoester and thioester in the S1C construct can be hydrolyzed by water. Furthermore, the thioester can be cleaved efficiently by the nucleophiles DTT and hydroxylamine, whereas the oxoester formed by the wild-type intein exhibits a higher stability under these conditions.
conditions, as proposed earlier for other inteins (9). Activation of the scissile peptide bond depends on the conserved block B/N3 His-90.

Because the most remarkable feature of this intein centers around formation of a linear oxoester that is transesterified into a thioester we had a closer look at other residues of the block B/N3 motif, which is involved in activation of the upstream
**Intein with Unprecedented Nucleophile Combination**

splice junction (14, 30, 34, 37–41). Instead of the consensus XTXXH motif, the GOS-TerL intein features the sequence CSXXH. Investigation of the aforementioned H90A mutation as a single substitution showed complete abrogation of protein splicing, whereas significant amounts of C-terminal cleavage products were formed (Fig. 7a). More surprisingly, S87A and S87T mutations also virtually completely inhibited splicing. In addition to C-terminal cleavage also, minor amounts of N-terminal cleavage could be observed. Thus, Ser-87 appeared essential for splicing with Ser1, but only important for the N-O

---

**TABLE 3**

Activity of the GOS-TerL intein with different mutations

<table>
<thead>
<tr>
<th>Int&lt;sup&gt;Thr&lt;/sup&gt;</th>
<th>Int&lt;sup&gt;Car&lt;/sup&gt;</th>
<th>Splicing&lt;sup&gt;a&lt;/sup&gt;</th>
<th>N-cleavage</th>
<th>C-cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>Wild-type</td>
<td>+ + +</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S1C</td>
<td>Wild-type</td>
<td>+ + +</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S1A</td>
<td>Wild-type</td>
<td>-</td>
<td>-</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>S87A</td>
<td>-</td>
<td>-</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>S87T</td>
<td>-</td>
<td>-</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>H90A</td>
<td>-</td>
<td>-</td>
<td>+ +</td>
</tr>
<tr>
<td>S1C</td>
<td>S87A</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>S1C</td>
<td>S87T</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>S1C</td>
<td>H90A</td>
<td>-</td>
<td>-</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>C86A</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Wild-type</td>
<td>C86S</td>
<td>-</td>
<td>-</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>C117A</td>
<td>+ + +</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wild-type</td>
<td>C86S, C117A</td>
<td>+ + +</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wild-type</td>
<td>D133A</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Wild-type</td>
<td>N152A</td>
<td>-</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>C + 1A</td>
<td>-</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>C + 1S</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>S1C</td>
<td>C + 1S</td>
<td>-</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>N152A, C + 1A</td>
<td>-</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>Wild-type</td>
<td>H90A, N152A, C + 1A</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S1C</td>
<td>H90A, N152A, C + 1A</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup> Symbols: -, indicates <10% activity; + indicates 10–30% activity; ++ indicates 30–60% activity; +++ indicates >60% activity.

---

**FIGURE 6. Analysis of the acyl-shift at the upstream splice junction of the GOS-TerL intein.** Shown are Commassie Brilliant Blue-stained SDS-PAGE gels. Reactions were performed at 30 °C. a, mutational analysis of the residue at position 1 the Int<sup>Thr</sup>-protein at 20 μM and Int<sup>Car</sup>-protein at 10 μM. b, time courses for the wild-type GOS-TerL intein and the S1C mutant. c, assessment of cleavage at the upstream scissile bond resulting in liberation of MBP. Wild-type (left) and S1C (right) Int<sup>Thr</sup> constructs (10 μM) were incubated with the N152A,C+1A mutant of the Int<sup>Thr</sup> protein (14.6 μM) and also probed with either 100 mM DTT, 2 mM tris(2-carboxyethyl)phosphine (TCEP), or 250 mM hydroxylamine (HA) for 20 h at pH 6, as indicated. Reactions were quenched by the addition of SDS buffer lacking DTT or β-mercaptoethanol and boiled for 10 min at 96 °C. Calculated molecular masses are: MBP-Int<sup> Thr</sup>-His<sub>6</sub> = 49.4 kDa; Int<sup> Thr</sup>-(N152A,C+1A)-Trx-His<sub>6</sub> = 27.1 kDa; MBP = 44.2 kDa; Int<sup> Thr</sup>-His<sub>6</sub> = 5.3 kDa. *, protein contamination.
where an internal cysteine side chain binds the N-extein as a thioester intermediate (11). Having already ruled out an essential role of Cys-86 for such a potential intermediate (see above), we also mutated the remaining cysteine at position 117. Fig. 7c shows that the intein with the C117A mutation exhibited full splicing activity. Also the double mutation C86S/C117A showed some remaining splicing activity, even though the stability of the protein against partial degradation was further reduced. Collectively, these results clearly show that the splicing pathway of the GOS-TerL intein does not use additional and atypical thioester intermediates.

**Mutational Analysis of the Block F/C2 Motif Reveals Strict Dependence of the GOS-TerL Intein on Cys**

The mutation of the conserved Asp in block F/C2 reduced splicing activity and uncoupled N- and C-terminal cleavage reactions (Fig. 7d) as previously reported for other inteins (42, 43). An individual mutation of the highly conserved Asn in the block G/C1 motif blocked splicing and all C-terminal activity, and also induced quantitative N-terminal cleavage (Fig. 7d). Together, these results are consistent with the established roles of the investigated residues in the canonical splicing pathway.

**FIGURE 7. Point mutations in conserved motifs and cysteine residues of the GOS-TerL intein.** Shown are Coomassie Brilliant Blue-stained SDS-PAGE gels. Reactions were performed at 30 °C in the presence of 2 mM DTT using a 2-fold excess of the IntN-protein (20 μM) over the IntC-protein (10 μM). a, analysis of amino acids in the block B/N3 motif. b, mutations in the block B/N3 motif in combination with a S1C mutation. c, analysis of internal cysteine residues. Mutation of Cys-86 resulted in reduced protein solubility. d, analysis of mutants in the block F/C2 and block G/C1 motifs. The calculated masses deviate only slightly from those given in Figs. 4 and 5 due to single point mutations. *, protein contamination.

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>List of IntN constructs used in this study</strong></td>
</tr>
<tr>
<td>Intein &amp; Name</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>GOS-TerL</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AceL-TerL</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>pep-C1</td>
</tr>
<tr>
<td>pep-C1S</td>
</tr>
</tbody>
</table>

* (MBP), maltose-binding protein with linker sequence; Fl, 5(6)-carboxyfluorescein; bold letters show additional amino acids.

**Intein with Unprecedented Nucleophile Combination**

The Ultrafast gp41-1 Intein and the Homologous Ace-TerL Inteins Are Inactive with a Ser1/Cys+1 Combination—The Ace-TerL intein is the closest characterized homolog of the GOS-TerL intein, however, it features a Cys/Cys combination at the two critical residues at the splice junctions. It also contains the uncommon serine residue in the block B/N3 motif, similar to Ser-87 in the GOS-TerL intein. Thus, we wondered whether the Ace-TerL intein may also support protein splicing with an artificially introduced Ser1. For this naturally split intein a semi-synthetic splice assay was previously established, in which the short IntN fragment (25 aa) was provided on a synthetic and fluorescently labeled peptide pep-C1 (25). We...
**Intein with Unprecedented Nucleophile Combination**

**FIGURE 8. Inactivity of the AceL-TerL intein with the Ser1/Cys+1 combination.** A synthetic Int^N^ peptide with 5(6)-carboxyfluorescein (Fl) in the Ex^N^ part (75 μl) was incubated with the Int^N^ construct (15 μl) of the AceL-TerL intein at 8°C. The Int^N^ peptide contained either the wild-type AceL-TerL sequence (pep-C1) or the C1S substitution (pep-C1S) as indicated. a, scheme of the reaction with the native nucleophile combination Cys/Cys+1. b, analysis by SDS-PAGE gels stained with Coomassie Brilliant Blue (top) and illuminated under UV light (bottom). Note that carboxyfluorescein causes a shift to top of the reaction with the native nucleophile combination Cys/Cys+1 or the C1S substitution (pep-C1S) as indicated. The native Int^N^ side chain of Cys+1 should be the better nucleophile than the synthetic Cys+1 combination should be observable more easily than with other inteins. Thus, the AceL-TerL intein does not support protein splicing with the Ser1/Cys+1 combination.

On a similar account, we chose the gp41-1 intein (26) as another test case for the unusual Ser1/Cys+1 combination (Fig. 9). This intein has a native Cys1/Ser+1 combination. One reason to test this intein was the CSXXH signature in its block B/N3 motif analogous to the GOS-TerL intein. Furthermore, the gp41-1 intein is the split intein with the highest kinetic rate reported so far (27). Thus, residual activity with a Ser1/Cys+1 combination should be observable more easily than with other inteins. First, we mutated the native Cys1 to serine. The mutation to generate the Cys1/Cys+1 combination resulted consistent with an earlier study (27). On the other hand, a S→C mutation which exhibited only slow C-cleavage (Fig. 9a, right panel) shows that no splice product formation could be observed. Significant amounts of N-terminal cleavage product indicated that a linear ester could be formed, but not be further processed in a directional manner (Fig. 9b, right panel). These findings underline once more how specialized even the most active inteins are toward their nucleophilic side chains at positions 1 and +1. Other factors must be important to make the Ser1/Cys+1 combination work.

The GOS-TerL Intein Tolerates a Truncation of Its Int^N^ Fragment and Is Capable of Cross-splicing with the AceL-TerL^N^ Fragment—We noticed that the sequence similarity between the Int^N^ pieces of the GOS-TerL and AceL-TerL inteins is very high in the last 11 amino acids of the shorter AceL-TerL intein, i.e. within the N2 motif. The GOS-TerL^N^ fragment has 8 or 9 amino acids ((S)FNERKFNE) as a C-terminal extension beyond this conserved block (Fig. 2). Furthermore, the sequence similarity is very low within the first 10 amino acids of the two Int^N^ pieces, including Ser1 of the GOS-TerL intein and Cys1 of the AceL-TerL intein. These observations led us to perform truncation experiments of the GOS^N^ fragment and cross-splicing assays between both inteins.

We could show that the model protein MBP-Int^N^Δ9aa-His$_S$ with a C terminally truncated variant of the GOS^N^ fragment was indeed capable of protein splicing with kinetics and yields indistinguishable from those of the wild-type construct (data not shown). Thus, the C-terminal amino acids of the GOS^N^ fragment seem to be superfluous at least under these experimental conditions. This finding also suggests that the split site of the GOS-TerL intein corresponds to that of the AceL-TerL intein, e.g. in the loop sequence before β-sheet 4 of the conserved HINT-fold (44, 45).

Next, we tested exchanging of the Int^N^ fragments of the GOS-TerL and AceL-TerL inteins in cross-splicing experiments. Very surprisingly, the GOS^C^ construct was also capable of splicing with the heterologous AceL^N^ peptide pep-C1 (Fig. 10a). Even more strikingly, the GOS^C^ could also form a splice product with the pep-C1S peptide (Fig. 10a). In contrast, the
AceLC protein could only splice with its corresponding AceLN partner peptide or protein (Fig. 10b), but with neither the full-length nor the C terminally truncated GOSN proteins (Fig. 10, c and d). The restricted activity of the AceL-TerL intein was independent from the nature of the residue at position 1, as AceLC was also incapable of splicing with the GOSN(S1C) or shortened GOSN(S1C, Δ9aa) mutants (Fig. 10, c and d). The promiscuity of the GOSC fragment toward both AceLN fragments is truly remarkable, given the diversity of the sequences within the first 10 amino acids that are expected to be a crucial part of the refolded intein complex. These results suggest that the exact sequence of the GOSN fragment is not important for the ability of the intein to splice with a Ser1 residue in general and with the unique Ser1/Cys1 combination in particular. Rather, the special features that enable this intein for this unusual mechanism should be located within the GOSC part.

**Conclusions**

The newly discovered GOS-TerL intein is the first and only intein described so far with the unexpected Ser1/Cys1 combination of residues at the two splice junctions. Also artificial introduction of this combination in other inteins was not reported yet to result in efficient protein splicing (17). As shown in this work, this also holds true for the gp41-1 and Ace-TerL inteins that naturally employ Cys1/Ser1 and Cys1/Cys1 combinations, respectively. These two inteins represent the fastest intein known and the closest homolog of the GOS-TerL intein, respectively.

Collectively, our mutational analysis of the GOS-TerL intein clearly supports the notion that its splicing pathway follows that of a typical class 1 intein in terms of the overall logic of the reactive intermediates and the participating key residues. Thus, the upstream exteinN-intein peptide bond is first rearranged into the oxoester, which is then converted into the thioester of the branched intermediate, before cleavage of the downstream intein-exteinC peptide bond irreversibly progresses the splicing mechanism beyond the potential equilibrium between the starting material and the two ester intermediates. Liberation of the N-terminal amine group of the exteinC also enables rearrangement into the new exteinN-exteinC peptide bond (Fig. 1b).

Why is the Ser1/Cys1 combination so unusual in the protein splicing pathway? One could speculate that in the evolution of inteins this combination was avoided because the shift from an oxoester as the first intermediate to a thioester as the second intermediate is an uphill reaction in a thermodynamic sense. The intein would have to catalyze two consecutive uphill steps starting from the peptide bond at the upstream splice junction as the ground state. In contrast, in all other previously observed combinations the transesterification step to the branched intermediate is either formally a downhill reaction (in Cys1/Ser1 and Cys1/Thr1 combinations) or at least a reaction on the same energetic level (in Cys1/Cys1, Ser1/Ser1, and Ser1/Thr1 combinations). Following the initial uphill reaction from the peptide bond to the linear oxoester or thioester intermediate, a downhill or equal level transesterification
reaction may be important to control the desired flux in the equilibrium of intermediates along the reaction coordinate of the protein splicing pathway. A second uphill reaction may be disfavored for this purpose and block splicing. Notably, also the splicing pathways of class 2 and class 3 inteins contain only one formal uphill reaction (11).

Of course, each intein may simply be highly specialized for the activation of its particular catalytic cysteine, serine, or threonine residues. It is conceivable that in an optimized local environment the reactivity and positioning of the catalytic side chains as well as the acyl intermediates may be tuned for each of the mentioned combinations to off-set the energetic differences. However, the very uneven distribution of combinations of the 1 and +1 residues in natural inteins as well as mutagenesis studies clearly indicate that the Ser1/Cys+1 combination is strongly disfavored. Thus, we propose that the GOS-TerL intein either modulates activities of the nucleophiles by an optimized microenvironment or employs very efficient strategies to irreversibly remove the branched thioester intermediate, i.e. by an efficiently coupled asparagine cyclization.

The GOS-TerL intein tolerates well the substitution of the catalytic Ser1 to cysteine (resulting in the Cys1/Cys+1 combination). In contrast, it is highly specialized on the Cys+1 nucleophile, because the Ser1/Ser+1 combination is inactive. Even the combination Cys1/Ser+1 is virtually inactive, although it is found in many other inteins.

Notably, mutations from Ser to Cys at both splice junctions have yielded active inteins, as shown here for the GOS-TerL and the gp41-1 inteins, respectively. Mutations from Cys to Ser/Thr have also been reported to work at the +1 position (46, 47), although they yield nearly inactive inteins in most cases (28). However, a Cys to Ser mutation at the 1 position that retained high splicing efficiency has not been reported to our knowledge. Interestingly, this mutation can give rise to high yields of N-terminal cleavage, as shown here, for example, by the gp41-1 intein, indicating that the N-O shift can still be performed. These findings point to the transesterification reaction being the critical step in artificial Ser1/Cys+1 or Ser1/Ser+1 combinations, either because of non-optimized electronic activation of the reaction partners or because of their slightly mispositioned arrangement in the active site. Alternatively, the transesterification step may occur in these cases, but the branched intermediate is hydrolyzed before asparagine cyclization can take place.

The amino acids in block B/N3 are important for the initial N-S or N-O shift. The His and Thr (or Ser) in the TXXH motif have been found to be critical or important in many inteins. Interestingly, in some cases one of these residues can be mutated to Ala without substantial loss of activity (34, 37). For the GOS-TerL intein with the native Ser1/Cys+1 combination both Ser-87 and His-90 were found to be indispensable. Interestingly, however, in the artificial Cys/Cys+1 combination, Ser-87 of block B/N3 could be mutated to Ala, whereas His-90 was still indispensable. This finding is in agreement with the idea that the Ser1-oxoester is chemically less activated than the Cys1-thioester. Therefore, formation or stabilization of the Ser1-oxoester should require more catalytic assistance from the intein. The Thr or Ser at the conserved position of the Ser-87 in the GOS-TerL intein have previously been suggested to help in distorting the scissile peptide bond at the upstream splice junction to facilitate the N-S acyl shift (41). A similar role in the N-O shift as well as a role in preparing the oxoester for the transesterification reaction is conceivable for the GOS-TerL intein.

Strikingly, the GOSC fragment forms a highly active intein with the AceLN fragment whose sequence of the N1 motif is substantially different. Also a C1S mutation in the AceLN2 fragment, which reconstitutes the Ser1/Cys+1 combination, yields an active AceLN2/GOSC hybrid intein system. Thus, the most important determinants for the usual splicing pathway of the GOS-TerL intein seem to be located in its IntC fragment. These remain to be identified in future studies to further unravel the enzymatic mechanism for the unprecedented sequence of acyl shifts.

The fact that the GOSN fragment is free of cysteine residues makes this intein an interesting candidate for certain biotechnological applications, such as protein labeling approaches (48–50). The small size of this fragment of 37 or 28 aa also calls for applications using semisynthetic protein trans-splicing (23, 25, 51, 52).

In conclusion, the GOS-TerL intein can master the unprecedented splicing pathway from peptide bond to oxoester to thioester, with seemingly two steps uphill with respect to the energy levels of the intermediates. Its splicing mechanism thus represents an intriguing diversion from the canonical pathway of class I inteins.

Author Contributions—A. L. B. and H. D. M. conceived the study, analyzed the data, and wrote the paper. A. L. B. performed all experiments except mass spectrometry.

Acknowledgments—We thank Shmuel Pietrokovski (Weizmann Institute of Science, Rehovot, Israel) and Julian Matern for helpful discussions, as well as Wolfgang Dörner and Julian Matern for support with mass spectrometry. We are grateful to Julia Diecker for excellent technical assistance.

References
Intein with Unprecedented Nucleophile Combination


