Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources[*]

Patricia Carvajal-Vallejos[‡1], Roser Pallissé[‡], Henning D. Mootz[§2], and Stefan R. Schmidt[‡]

From [‡]ERA BIOTECH®, 08028 Barcelona, Spain and the [§]Institute of Biochemistry, University of Münster, 48149 Münster, Germany

Background: Hypothetical natural split inteins from environmental metagenomic sources were evaluated experimentally for the first time.

Results: Four inteins showed the highest known reaction rates and efficiencies for protein *trans*-splicing and could be used efficiently for C-cleavage.

Conclusion: These inteins pushed the catalytic limit of protein *trans*-splicing, although this was unpredictable from their primary sequence.

Significance: These inteins hold great potential as versatile protein engineering tools.

Inteins excise themselves out of precursor proteins by the protein splicing reaction and have emerged as valuable protein engineering tools in numerous and diverse biotechnological applications. Split inteins have recently attracted particular interest because of the opportunities associated with generating a protein from two separate polypeptides and with *trans*-cleavage applications made possible by split intein mutants. However, natural split inteins are rare and differ greatly in their usefulness with regard to the achievable rates and yields. Here we report the first functional characterization of new split inteins previously identified by bioinformatics from metagenomic sources. The N- and C-terminal fragments of the four inteins gp41-1, gp41-8, NrdJ-1, and IMPDH-1 were prepared as fusion constructs with model proteins. Upon incubation of complementary pairs, we observed *trans*-splicing reactions with unprecedented rates and yields for all four inteins. Furthermore, no side reactions were detectable, and the precursor constructs were consumed virtually quantitatively. The rate for the gp41-1 intein, the most active intein on all accounts, was $k = 1.8 \pm 0.5 \times 10^{-1} \, \text{s}^{-1}$, which is ~10-fold faster than the rate reported for the *Npu* DnaE intein and gives rise to completed reactions within 20–30 s. No cross-reactivity in exogenous combinations was observed. Using C1A mutants, all inteins were efficient in the C-terminal cleavage reaction, albeit at lower rates. C-terminal cleavage could be performed under a wide range of reaction conditions and also in the absence of native extein residues flanking the intein. Thus, these inteins hold great potential for splicing and cleavage applications.

Inteins, a combination of the words "internal protein" (1), were first discovered in 1990 when it was observed that the

translation of the *VMA1* gene of *Saccharomyces cerevisiae* did not result in the predicted 120-kDa protein, but a much shorter polypeptide chain of 69 kDa (2). To obtain the mature 69-kDa protein, an internal part of the larger precursor had to be excised and the flanking regions, called N- and C-extein, rejoined in an autocatalytic post-translational process termed protein splicing. Furthermore, the excised intein fragment contained a nonessential part exhibiting homing endonuclease activity that could be deleted without affecting the overall ability of the mini-intein to cleave and splice (3). More recently, inteins without the endonuclease domain have also been found (4). Since that time, almost 500 other inteins have been identified in the three kingdoms of organisms (archaea, bacteria, and eukaryotes), and their sequences have been collected in a dedicated database (5). Overall, four short conserved regions on the amino terminus and two conserved domains on the carboxyl terminus could be identified that can be used to search for novel inteins (1, 6). Besides the inteins that are encoded in a single gene, a few split inteins were also discovered and characterized biochemically that are transcribed and translated from two separate genes and lack the endonuclease domain. These include a number of alleles inserted in the cyanobacterial DnaE gene encoding the α subunit of DNA polymerase (7), such as *Npu* DnaE (8), *Ssp* DnaE (9), and *Anabaena* sp. DnaE inteins (10) as well as the archaeon *Neq* polymerase intein (11). As shown in Fig. 1*A*, the translation of both fragments reconstitutes the catalytic activity of protein *trans*-splicing to form the mature protein (12). Split inteins can also be generated artificially by dividing inteins into two or three fragments (13). In the case of the *Ssp* DnaB intein, the N-terminal intein fragment (Int[N])[3] could be as short as 11 amino acids (14), whereas 6 amino acids were sufficient for the Int[C] fragment of the *Ssp* GyrB intein (15).

Because inteins are in principle tolerant with regard to heterologous sequences at their flanking extein regions, they can be inserted into other proteins. This opens the door for many

---

[1] To whom correspondence may be addressed: ERA BIOTECH®, Baldiri i Reixac 15-21, 08028 Barcelona, Spain. Tel.: 34-934-034-771; Fax: 34-934-034-772; E-mail: pcarvajal@erabiotech.com.

[2] To whom correspondence may be addressed: Institute of Biochemistry, University of Muenster, Wilhelm-Klemm-Str. 2, 48149 Münster, Germany. Tel.: 49-251-83-33-005; Fax: 49-251-83-33-007; E-mail: henning.mootz@uni-muenster.de.

[3] The abbreviations used are: Int[N], intein-N; Int[C], intein-C; 5aa[N], native extein-N; 5aa[C], native extein-C; ST, Strep-tag II; gpD, bacteriophage λ head protein D; Trx, thioredoxin; H[6], His tag; SP, splice product; IMPDH, inosine-5′-monophosphate dehydrogenase; Bis-Tris, 2-(bis(2-hydroxyethyl)amino)-2-(hydroxymethyl)propane-1,3-diol.

applications, for instance, to generate cyclic peptides (16), to generate protein thioesters (17), to label proteins (18, 19), and probably most frequently, to cleave off affinity tags for protein purification (20). Particularly to optimize cleavage efficiency, inteins have been mutagenized to become inducible by pH (21), reducing agents (22), or temperature (23) and to separate cleavage activity from splicing. Interestingly, inteins can be engineered to cut either at the amino terminus (22) or at the carboxyl terminus (24).

Many of the artificial split inteins require refolding or show low solubility in fusion constructs, which limits their application. However, split inteins are especially attractive for biotechnological applications because they have no premature activity when expressed separately and they are much smaller than normal inteins and thus potentially have less negative impact on the expression level of the fused target protein (25). The ideal intein for some biotechnical applications would therefore be a split intein that is highly active without the need of refolding. Currently, the *Npu* DnaE intein appears to be the most advantageous split intein as it displays superior splicing kinetics and high efficiency (8, 26).

After the first isolation of naturally occurring split inteins, efforts were undertaken to find more native split inteins. Most recently, a search in Global Ocean Sampling (GOS) environmental metagenomic sequence data was performed to find fractured genes that contain novel split inteins (27). These fractured genes code for potentially essential cellular proteins, and in the case of four insertion sites (gp41 DNA helicase, inosine-5′-monophosphate dehydrogenase (IMPDH), ribonucleotide reductase catalytic subunit NrdJ, and DnaE polymerase II subunit α), the complete loci including the split inteins could be assembled. However, it remained unclear whether the inteins were active in protein *trans*-splicing.

The purpose of this study was to functionally characterize four representatives of these novel hypothetical split inteins (gp41-1, gp41-8, NrdJ-1, and IMPDH-1) in the protein *trans*-splicing reaction and to perform mutagenesis to evaluate their efficiency and applicability for *trans*-cleavage. We found that these inteins exhibit an unprecedented activity in terms of reaction speed and efficiency.

EXPERIMENTAL PROCEDURES

Plasmid Construction—Gene sequences and specific details of novel split inteins gp41-1, gp41-8, NrdJ-1, and IMPDH-1 are already described elsewhere (27). Gene synthesis of split intein fragments was performed by Biomatik (Cambridge, Canada), including flanking restriction sites for easy subcloning into the previously described pAU08 and pVS01 plasmids (26). Construction of the expression plasmids encoding the Int^N fragments of gp41-1, gp41-8, NrdJ-1, and IMPDH-1 together with their 5 native extein residues ($5aa^N$) was accomplished by BamHI and HindIII restriction digestion (Fermentas, St. Leon-Rot, Germany) and by their subcloning into pAU08 to replace the *Npu* DnaE^N; an ST-gpD sequence is encoded upstream of the Int^N fragments. The four generated expression plasmids were named as follows: gp41-1^N, gp41-8^N, NrdJ-1^N, and IMPDH-1^N. Plasmids encoding the Int^C split intein fragments of gp41-1, gp41-8, NrdJ-1, and IMPDH-1 together with their 5

A

B

ST-**gpD**-5aa^N-**Int^N**-H$_6$ +   **Int^C**-5aa^C-**Trx**-H$_6$

proteins 1-4 and 11-14(C1A)   **proteins 6-9 and 15-18(w/o 5aa^C)**

protein *trans*-splicing   protein *trans*-cleavage (C1A)

ST-**gpD**-5aa^N-5aa^C-**Trx**-H$_6$ 5aa^C-**Trx**-H$_6$

splice product (SP)   cleavage product (Trx)

+   +

Int^N-H$_6$ + **Int^C**   ST-**gpD**-5aa^N-**Int^N**-H$_6$ + **Int^C**
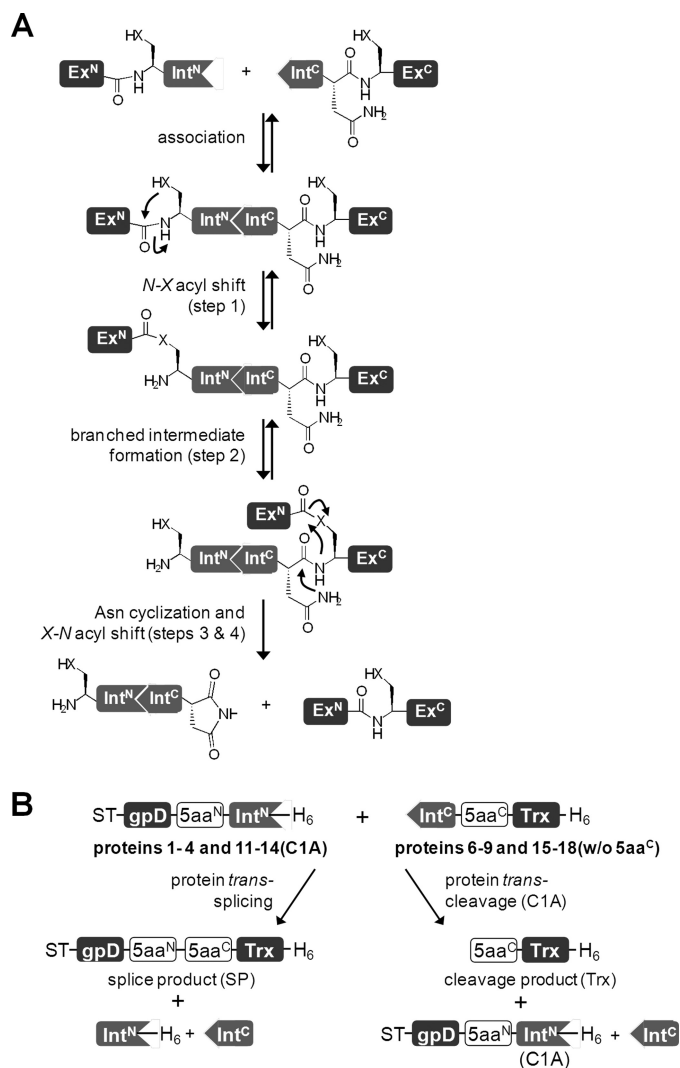   (C1A)

FIGURE 1. **Schemes of the protein *trans*-splicing pathway and of the constructs used in this study.** *A*, following intein fragment association, an N–S or N–O acyl shift forms a thioester or oxoester bond at the N-extein/intein junction. This reactive intermediate is attacked in a *trans*(thio)esterification by the side chain sulfhydryl or hydroxyl group of the first residue in the C-extein, which can be Cys, Ser, or Thr, to give a branched intermediate. The cyclization of the conserved Asn residue at the C terminus of the intein releases the intein. Finally, the (thio)ester bond between the exteins rearranges to a peptide bond by a spontaneous S–N or O–N acyl shift. In general, *X* can be sulfur or oxygen. Note that all four new inteins characterized here employ a Cys-1 and a Ser+1 as catalytic residues at the splice junctions. *B*, shown are the intein fusion proteins used in this study and the products of the protein *trans*-splicing (*left panel*) and the *trans*-cleavage reactions (*right panel*). ST-gpD-5aa^N is the Extein^N and 5aa^C-Trx-H$_6$ is the Extein^C. 5aa$^{(N/C)}$ represent the 5 native extein amino acids of the respective intein flanking the intein domain N- or C-terminally.

native extein residues ($5aa^C$) were obtained by KpnI and NdeI (Fermentas, St. Leon-Rot, Germany) restriction digestion and by their subcloning into pVS01 to replace the *Npu* DnaE^C; a Trx-His tag sequence is downstream of the Int^C fragments. The four generated expression plasmids were named as follows: gp41-1^C, gp41-8^C, NrdJ-1^C, and IMPDH-1^C.

Single-point mutation to allow C-terminal cleavage was performed in constructs containing the Int^N fragments. A C1A mutation for each split intein was performed by PCR using specific primers to give gp41-1$^{N(C1A)}$, gp41-8$^{N(C1A)}$, NrdJ-1$^{N(C1A)}$, and IMPDH-1$^{N(C1A)}$. Synthetic oligonucleotides were obtained from Thermo Scientific (Ulm, Germany).

TABLE 1
Protein constructs used in this study

Protein	Abbreviation	Protein sequence[a]	Molecular mass
			kDa
1	gp41-1[N]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*TRSGY*]-gp41-1[N]-GG<u>HHHHHH</u>	24.2
2	gp41-8[N]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*SQLNR*]-gp41-8[N]-GG<u>HHHHHH</u>	24.2
3	NrdJ-1[N]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*GTNPC*]-NrdJ-1[N]-GG<u>HHHHHH</u>	26.2
4	IMPDH-1[N]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*GIGGG*]-IMPDH-1[N]-GG<u>HHHHHH</u>	25.8
5[b]	*Npu* DnaE[N]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*-*Npu* DnaE[N]	24.7
6	gp41-1[C]	gp41-1[C]-[*SSSDV*]*GT*-Trx-EFRS<u>HHHHHH</u>	18.8
7	gp41-8[C]	gp41-8[C]-[*SAVEE*]*GT*-Trx-EFRS<u>HHHHHH</u>	19.1
8	NrdJ-1[C]	NrdJ-1[C]-[*SEIVL*]*GT*-Trx-EFRS<u>HHHHHH</u>	18.6
9	IMPDH-1[C]	IMPDH-1[C]-[*SICST*]*GT*-Trx-EFRS<u>HHHHHH</u>	18.6
10[b]	*Npu* DnaE[C]	*Npu* DnaE[C]-[*CFN*]*GT*-Trx-EFRS<u>HHHHHH</u>	17.9
11	gp41-1[N (C1A)]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*TRSGY*]-gp41-1[N (C1A)]-GG<u>HHHHHH</u>	24.2
12	gp41-8[N (C1A)]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*SQLNR*]-gp41-8[N (C1A)]-GG<u>HHHHHH</u>	24.2
13	NrdJ-1[N (C1A)]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*GTNPC*]-NrdJ-1[N (C1A)]-GG<u>HHHHHH</u>	26.2
14	IMPDH-1[N (C1A)]	MAS<u>WSHPQFEKAS</u>-gpD-*GS*[*GIGGG*]-IMPDH-1[N (C1A)]-GG<u>HHHHHH</u>	25.8
15	gp41-1[C (Δext)]	gp41-1[C]-*GT*-Trx-EFRS<u>HHHHHH</u>	18.3
16	gp41-8[C (Δext)]	gp41-8[C]-*GT*-Trx-EFRS<u>HHHHHH</u>	18.6
17	NrdJ-1[C (Δext)]	NrdJ-1[C]-*GT*-Trx-EFRS<u>HHHHHH</u>	18.2
18	IMPDH-1[C (Δext)]	IMPDH-1[C]-*GT*-Trx-EFRS<u>HHHHHH</u>	18.1

[a] Native extein sequences are in brackets, StreptagII- and His-tag are underlined.
[b] These constructs were reported in Ref. 26.

To permit a clean C-terminal cleavage, 5aa[C] residues placed between the Int[C] and the thioredoxin (Trx) encoding gene sequence were removed by PCR using specific primers to give gp41-1[C(Δext)], gp41-8[C(Δext)], NrdJ-1[C(Δext)], and IMPDH-1[C(Δext)]. All plasmids were verified by DNA sequencing (Macrogen). Table 1 gives an overview of all constructs used in this study and their numbering.

Expression of the Recombinant Intein Fusion Proteins—All generated plasmids were used for transformation of *Escherichia coli* BL21 (DE3) cells (Stratagene). Cells were grown at 37 °C and 250 rpm in shake flasks containing 600 ml of LB medium supplemented with the corresponding plasmid maintenance antibiotics (50 μg/ml kanamycin or 100 μg/ml ampicillin), until $A_{600\,nm}$ reached ~0.6. Induction of gene expression was then triggered by the addition of 0.4 mM isopropyl-β-D-thiogalacto-pyranoside for the Int[N] constructs or by the addition of 0.02% L-arabinose for the Int[C] constructs; in both cases, the temperature was lowered to 28 °C. After a 5-h induction, cells were harvested by centrifugation for 15 min at 5000 rpm, and the cell pellet was stored at −80 °C.

Split Intein Fusion Protein Purification—Protein purification was always performed under native conditions. For the Int[N] constructs, the cell pellet was resuspended in buffer containing 100 mM Tris/HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA. For the Int[C] constructs, cells were resuspended in buffer containing 50 mM Tris/HCl (pH 8.0), 300 mM NaCl, and 20 mM imidazole. Cells were sonicated on ice for 20 min with a 30% pulsed activity cycle (Branson 250 Sonifier) and centrifuged for 30 min at 34,500 × *g* at 4 °C. The soluble fraction of Int[N] fusion proteins was purified on Strep-Tactin columns (IBA, Göttingen, Germany), whereas soluble Int[C] constructs were purified on Ni²⁺-nitrilotriacetic acid (IBA) columns following the manufacturer's instructions.

Eluted fractions containing the purified fusion proteins were pooled, dialyzed against splicing buffer (50 mM Tris/HCl at pH 7.0, 300 mM NaCl, 1 mM EDTA, 10% (v/v) glycerol) in the presence of 2 mM DTT, and stored in aliquots at −80 °C. Protein concentration was determined using the calculated absorbance at 280 nm.

trans-Splicing and C-terminal trans-Cleavage Assays—In vitro reactions were performed basically as previously described (12, 26). In brief, purified N- and C-fusion proteins were shortly preincubated separately in the corresponding testing conditions. Protein splicing and C-terminal cleavage reactions were started by mixing complementary N- and C-fusion proteins in splicing buffer or different buffer conditions as indicated at equimolar concentrations of 5 μM and incubating at different temperatures. Aliquots were removed at specific time intervals, and the reaction was stopped by the addition of SDS-PAGE sample buffer containing 8% SDS (w/v) and 20% β-mercaptoethanol (v/v) followed by boiling. For reactions carried out in the presence of different urea concentrations or at different pH values, proteins were first dialyzed against the corresponding buffer. All reactions were monitored by SDS-PAGE (4–12% Bis-Tris gels from Novex, Invitrogen) followed by Coomassie Brilliant Blue (Sigma) staining.

Kinetic Analysis of trans-Splicing and C-terminal trans-Cleavage—Relative intensities of protein bands were determined densitometrically using the Quantity One (Bio-Rad) program. The different splicing/cleavage products were normalized according to their corresponding molecular weight. The percentage of protein splicing was calculated from the ratio of the splice product and the most consumed precursor. Constant rates (k_{obs}) were determined using the GraFit software (Erithacus, Surrey, UK), by fitting the data to the equation $P = P_0(1 - e^{-kt})$, where P is the percentage of splice product formation at time t, P_0 is the maximum percentage of spliced product obtained (yield), and k is the observed rate (12). All reactions were treated as irreversible, presteady-state first-order processes under the assumption that after a fast association of the two complementary fragments, protein *trans*-splicing or self-cleavage proceeded like a monomolecular reaction.

RESULTS

Generation and Purification of Split Intein Fusion Constructs—The four split inteins gp41-1, gp41-8, NrdJ-1, and IMPDH-1 were prepared as Int[N] and Int[C] fusion partners for the *in vitro* analysis. As shown in Fig. 1*B*, all the fusion proteins contained

the sequence of bacteriophage λ head protein D (gpD) and thioredoxin (Trx) as N- and C-exteins, respectively. Strep-tag II (ST) and/or His tag (H_6) were added for purification purposes. Five native extein amino acids (5aa) were kept upstream and downstream of the Int^N and Int^C fragments. Additionally, the non-native amino acid sequence GS upstream of the native $Extein^N$ residues and GT downstream of the native $Extein^C$ residues were added due to cloning reasons (Table 1). All proteins were expressed in soluble form in *E. coli* and purified under native conditions. The *Npu* DnaE intein, considered to have the highest splicing rate yet described, was used as a control in this study with previously described constructs (26) that exhibit equivalent fusion protein design (Fig. 1*B* and Table 1).

Time-dependent in Vitro trans-Splicing Analysis—Protein splicing is a self-processing pathway mediated by the intein domain. In the most common class I inteins (28), four successive nucleophilic displacement reactions that directly involve the side chain functional groups of three of the four splice junction residues lead to intein excision with concomitant ligation of the N- and C-terminal exteins. In protein *trans*-splicing, this pathway is preceded by intein fragment association, as depicted in Fig. 1*A*. Purified Int^N fusions (proteins 1–5 in Table 1) and Int^C fusions (proteins 6–10 in Table 1) were mixed in pairwise combinations in a 1:1 stoichiometry. For each intein, the spontaneous *trans*-splicing reaction was monitored at different temperatures for the formation of the splice product (SP) ST-gpD-Trx-H_6 and the consumption of the Int^N and Int^C precursors in SDS-PAGE gels. Also, the formation of the excised Int^N and Int^C fragments could be observed (compare Fig. 1*B*). The identity of the expected SP sequence was confirmed in each case by immunoblotting with anti-Strep-tag II and anti-Trx (data not shown). Moreover, the SP and also the trypsin fragment covering the new peptide bond linking the exteins were identified by LC-MS/MS (data not shown).

To determine the kinetics of the reactions, aliquots were removed at the specified time points and immediately quenched by mixing with SDS sample buffer and boiling. Surprisingly, the four new split inteins showed extremely fast *trans*-splicing reactions. In particular, the gp41-1 intein exhibited outstanding splice kinetics and yields with optimal performance at 45 °C. As shown in Fig. 2*A*, already within the first seconds of the reaction, the two precursor proteins disappeared almost completely (*bands 1* and *6*), whereas the SP and the Int^N and Int^C halves were formed. The other inteins gp41-8, NrdJ-1, and IMPDH-1 also spliced exceptionally fast and efficiently. After 1 min of incubation at their optimal temperature (37 °C), the respective precursors (Fig. 2*B*, *bands 2* and *7*, *3* and *8*, and *4* and *9*) were nearly depleted, and the SP as well as the Int^N and Int^C halves were detected as the predominant bands in the SDS-PAGE gels. For comparison, the previously described *Npu* DnaE intein was examined at its optimal temperature of 37 °C (26). Fig. 2*C* shows that SP formation occurred at a lower rate than seen for the new split inteins. For instance, after 1 min, the precursors are converted to the products to only about 50%.

For a more quantitative analysis, rates (*k*) and yields (in percentage) from the time course experiments were determined under different temperatures (Fig. 3). Rates were fitted to an exponential first-order equation with the assumption that
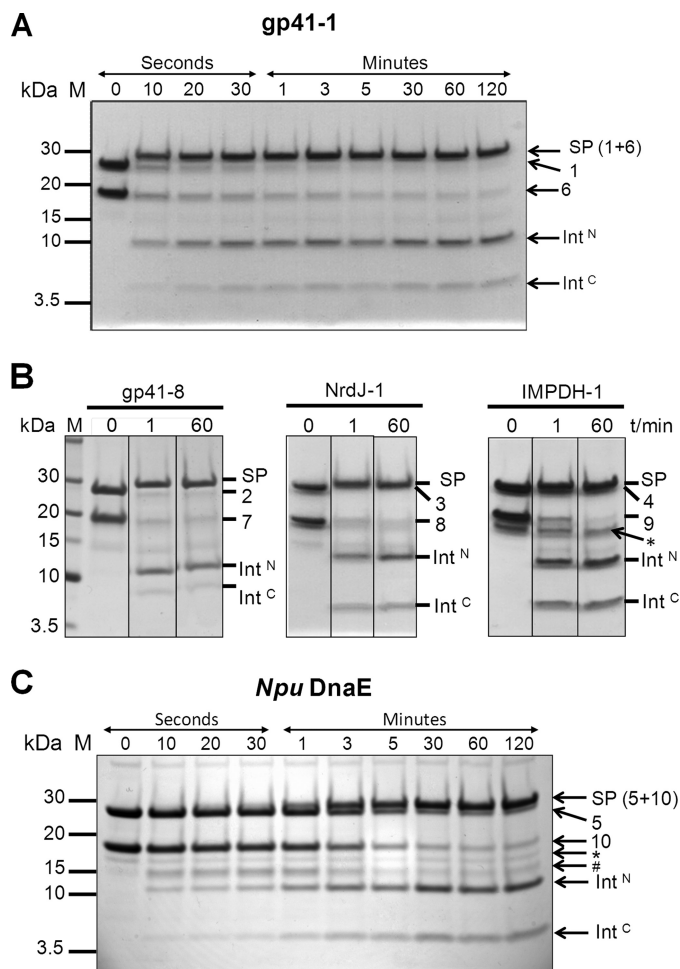

FIGURE 2. Analyses of protein *trans*-splicing reactions. Shown are Coomassie Brilliant Blue-stained SDS-gels (4–12% acrylamide). In all cases, the precursor proteins (indicated by *numbers*) were converted into the SP and the excised intein fragments Int^N and Int^C. Samples at $t = 0$ min are precursor proteins that were boiled before mixing. *A*, time-dependent analysis of the gp41-1 intein reaction performed at 45 °C. *M* denotes molecular marker. *B*, analyses of the reactions involving the gp41-8, NrdJ-1m and IMPDH-1 inteins performed at 37 °C. The band marked with an *asterisk* is a protein contamination. *C*, time-dependent analysis of the *Npu* DnaE intein reaction performed at 37 °C. The band marked with an *asterisk* denotes a protein impurity, whereas the band marked with a *number sign* corresponds to the N-terminal cleavage product (ST-gpD) resulting from cleavage of the linear thioester by β-mercaptoethanol in the SDS-sample buffer used to quench the reaction (as observed previously in Ref. 26).

intein fragment association is fast when compared with protein *trans*-splicing (29). All results of this comprehensive analysis are listed in Table 2. For the *Npu* DnaE intein, the rate of splice product formation $k = (1.4 \pm 0.3) \times 10^{-2}\,\mathrm{s^{-1}}$ corresponded to a $t_{1/2} = 50$ s at its optimal temperature of 37 °C (Fig. 3*A*), whereas 80–90% yield was observed under these conditions. These findings are in good agreement with the previous study (26). Fig. 3*A* shows that the gp41-1 intein was active over a wide range of temperatures and exhibited its highest rate and yield at 45 °C, with an unprecedented $k = (1.8 \pm 0.5) \times 10^{-1}\,\mathrm{s^{-1}}$, corresponding to a $t_{1/2} = 3.8$ s. Yields of 85–95% were reached with no apparent formation of side products by N- or C-terminal cleavage (see also Fig. 2*A*). Fig. 3*B* depicts detailed time courses of gp41-1 SP formation at different temperatures. This intein is thus ~13-fold times faster than the *Npu* DnaE intein at its optimal temperature and ~10-fold faster at 37 °C.
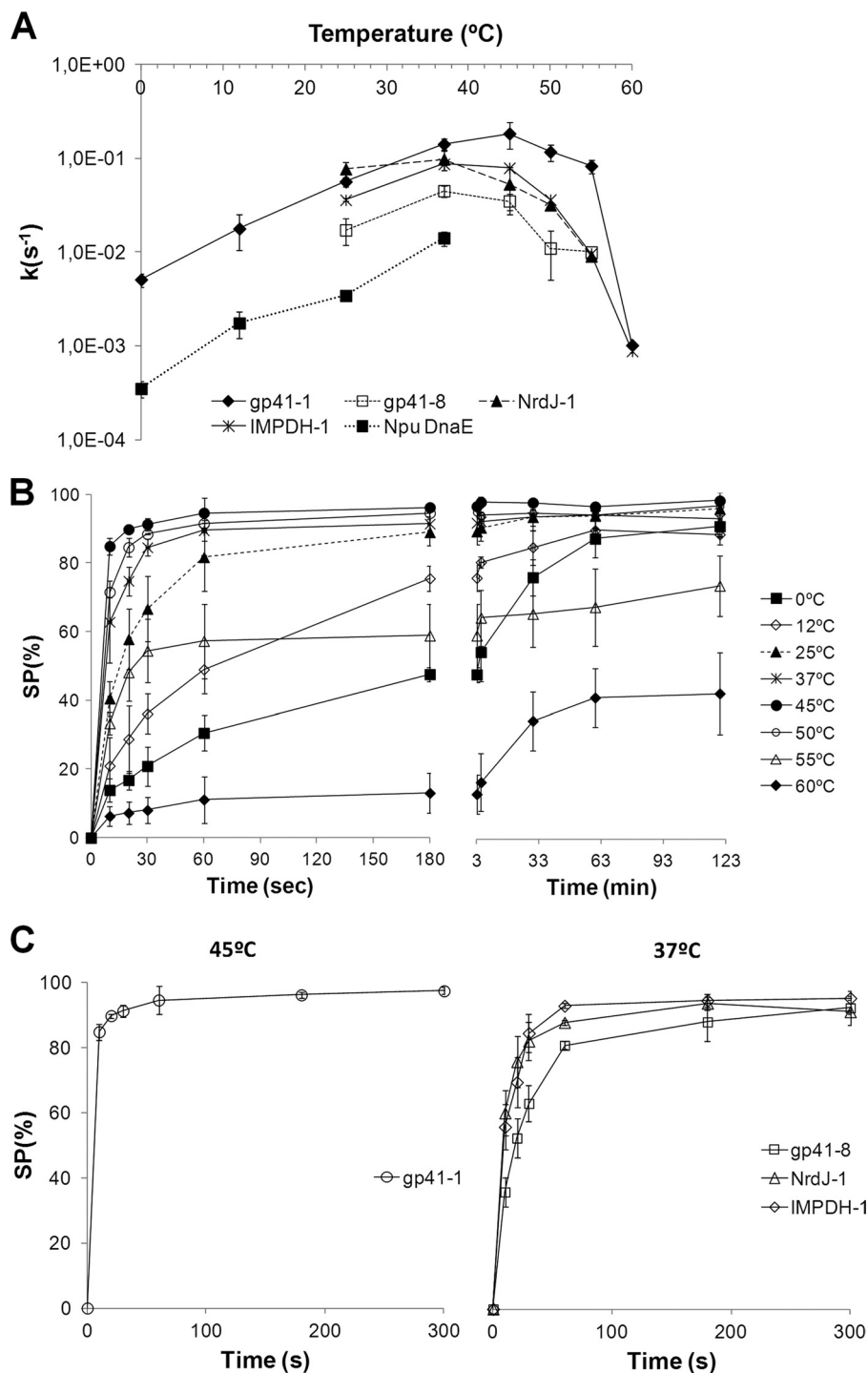
FIGURE 3. **Time courses and temperature dependence of protein *trans*-splicing reactions.** *A*, shown are constant rates as a function of temperature. Data for the *Npu* DnaE intein are included for comparison. *B*, time courses of SP formation catalyzed by the gp41-1 intein at different temperatures. *C*, time courses of splice product formation for the gp41-1 as well as gp41-8, NrdJ-1, and IMPDH-1 inteins at their optimal temperature as indicated and pH 7. *Error bars* in *panels A–C* indicate S.D.

The optimal temperature for the other three inteins was 37 °C (Fig. 3, *A* and *C*, *right panel*), at which the gp41-8 intein displayed a $k = (4.5 \pm 0.6) \times 10^{-2}\,\text{s}^{-1}$ with a calculated $t_{1/2} = 15\,\text{s}$ (85–95% yield). The NrdJ-1 intein had a $k = (9.8 \pm 2.3) \times 10^{-2}\,\text{s}^{-1}$ with $t_{1/2} = 7\,\text{s}$ (85–95% yield), and the IMPDH-1 intein had a $k = (8.7 \pm 3.2) \times 10^{-2}\,\text{s}^{-1}$ with $t_{1/2} = 8\,\text{s}$ (90–95% yield) (Table 2). These inteins are thus ~3-, 7- and 6-fold faster than the *Npu* DnaE. In Fig. 3*C*, the *left panel* shows the gp41-1 intein

at its optimal temperature for comparison. Considering their highest rates, these inteins can be arranged in a descending order as follows: gp41-1 > NrdJ-1 > IMPDH-1 > gp41-8 > *Npu* DnaE.

Orthogonality of the Split Inteins—To determine the interaction specificity, fusion constructs with the N- and C-terminal halves of the four new split inteins were mixed in all possible 16 combinations, resulting in four endogenous (wild type) and 12

TABLE 2

Apparent first order rate constants and yields of *trans*-splicing reactions - temperature dependence

NA = Not Active; ND = not determined; k = constant rate; Y = yield. Observed rates and yields were calculated as described under "Experimental Procedures." The time-courses for protein *trans*-splicing were monitored for 2 h and were started upon the addition of the C-domain to the N-domain (each at 5 μM concentration).

Temperature (°C)	gp41-1(1 + 6)		gp41-8(2 + 7)		NrdJ-1(3 + 8)		IMPDH-1(4 + 9)		*Npu* DnaE (5 + 10)	
	k	Y	k	Y	k	Y	k	Y	k	Y
	s^{-1}	%	s^{-1}	%	s^{-1}	%	s^{-1}	%	s^{-1}	%
0	$(5.1 \pm 0.8) \times 10^{-3}$	75–85							$(3.5 \pm 0.7) \times 10^{-4}$	75–85
12	$(1.8 \pm 0.7) \times 10^{-2}$	80–90							$(1.8 \pm 0.5) \times 10^{-3}$	75–85
25	$(5.7 \pm 0.6) \times 10^{-2}$	80–90	$(1.7 \pm 0.5) \times 10^{-2}$	85–95	$(7.7 \pm 1.5) \times 10^{-2}$	85–95	$(3.6 \pm 1.4) \times 10^{-2}$	90–95	$(3.4 \pm 0.3) \times 10^{-3}$	75–85
37	$(1.4 \pm 0.2) \times 10^{-1}$	90–95	$(4.5 \pm 0.6) \times 10^{-2}$	85–95	$(9.8 \pm 2.3) \times 10^{-2}$	85–95	$(8.7 \pm 3.2) \times 10^{-2}$	90–95	$(1.4 \pm 0.3) \times 10^{-2}$	80–90
45	$(1.8 \pm 0.5) \times 10^{-1}$	85–95	$(3.5 \pm 0.7) \times 10^{-2}$	75–85	$(5.3 \pm 2.8) \times 10^{-2}$	50–60	$(7.9 \pm 1.9) \times 10^{-2}$	90–95		
50	$(1.2 \pm 0.2) \times 10^{-1}$	85–95	$(1.1 \pm 0.6) \times 10^{-2}$	35–45	$(3.2 \pm ND) \times 10^{-2}$	30–40	$(3.6 \pm ND) \times 10^{-2}$	85–95		
55	$(8.3 \pm 1.3) \times 10^{-2}$	65–75	$(1.0 \pm ND) \times 10^{-2}$	30–35	$(9.0 \pm ND) \times 10^{-3}$	30–35	$(9.7 \pm ND) \times 10^{-3}$	80–90		
60	$(1.0 \pm 0.2) \times 10^{-3}$	40–50	NA		NA		$(8.8 \pm ND) \times 10^{-4}$	20–30		

TABLE 3

Effect of pH and urea on gp41-1 *trans*-splicing at 25 °C

ND = not determined; k = constant rate; Y = yield. $t_{1/2}$, half-life time. Observed rates and yields were calculated as described under "Experimental Procedures." The time-courses for protein *trans*-splicing were monitored for 2 h, and were started upon the addition of the C-domain to the N-domain (each at 5 μM concentration).

pH	Urea	gp41-1 (1 + 6)		
		k	$t_{1/2}$	Y
	M	s^{-1}		%
6	0	$(4.0 \pm 0.2) \times 10^{-2}$	17 s	80–90
8	0	$(5.4 \pm 0.2) \times 10^{-2}$	13 s	80–90
9	0	$(4.1 \pm 0.8) \times 10^{-2}$	17 s	80–90
10	0	$(4.5 \pm 1.2) \times 10^{-2}$	16 s	80–90
7	4	$(1.0 \pm ND) \times 10^{-3}$	11 min	80–85[a]
7	6	$(2.0 \pm ND) \times 10^{-5}$	577 min	30–35[a]

[a] 24-h time course.

exogenous combinations. Protein splicing was only detected in the endogenous combinations, showing that the inteins are orthogonal (data not shown).

Effect of Urea and pH on gp41-1 trans-Splicing—We tested the effect of a chaotropic salt and different pH conditions (pH 6–10) on *trans*-splicing of the gp41-1 intein (Table 3). Prior to mixing of the complementary intein constructs, the samples were equilibrated in the corresponding buffer conditions as indicated. gp41-1 intein activity remained very high in the presence of 4 M urea ($k = 1.0 \times 10^{-3}$ s^{-1}; yield = 80–85%), but decayed 50-fold with 6 M urea. There was no activity detected at 8 M urea. The gp41-1 intein was robust toward the pH variation, and the rates and yields remained similar from pH 6–10 (Table 3).

C-cleavage Induction with Point Mutation of Conserved Residues—Cleavage mediated by inteins is attractive for some biotechnological applications, *e.g.* bioprocess operations to release the protein of interest from a fusion partner. For this application, the intein must be altered to avoid *trans*-splicing and to effect a controllable cleavage. The mutation of the first Cys residue of an intein typically abolishes step 1 (the N–S acyl shift), step 2 (formation of the branched intermediate by attack of the first residue in the C-extein), and step 4 (rearrangement to a peptide bond by a spontaneous O–N acyl shift) of the splicing reaction (Fig. 1*A*). In such mutated inteins, step 3 (Asn cyclization) may still occur, leading to C-terminal cleavage of the C-extein from the N-extein/intein portion.

To assess the potential of the new split inteins as protein cleavage tools, we mutated the Cys-1 to Ala at the N-terminal splice junction of the four inteins, gp41-1, gp41-8, NrdJ-1, and IMPDH-1 (proteins 11–14; Fig. 1*B* and Table 1). The purified mutant fusion constructs were incubated with the previously

used C-terminal constructs 6–9, containing Trx as the product of interest and the 5 native C-extein residues flanking the intein (Fig. 1*B* and Table 1). C-cleavage activity at different temperatures (0, 25, and 37 °C) was determined by quantifying the protein bands by SDS-PAGE as described before.

In all the cases, C-cleavage activity with good (about 40–50%) to excellent (about 85–95%) yield was detected (Table 4), albeit at lower rates than observed for the *trans*-splicing reactions (Table 2). The corresponding SDS-PAGE gels are shown in Fig. 4, *A–D, left panels*. In each case, the bands of the precursor protein (*bands 6, 7, 8,* and *9* respectively) were cleaved into two products: the 5aa^C-Trx protein and the respective Int^C fragment. At 37 °C, the gp41-1 intein exhibited the highest C-cleavage rate among the new inteins $k = (2.2 \pm 0.7) \times 10^{-3}$ s^{-1}, corresponding to a calculated $t_{1/2}$ ~5 min with 85–95% yield followed by the IMPDH-1 intein $k = (7.7 \pm 1.6) \times 10^{-4}$ s^{-1}, $t_{1/2}$ ~15 min (85–95% yield), the NrdJ-1 intein $(2.4 \pm 0.1) \times 10^{-4}$ s^{-1}, $t_{1/2}$ ~48 min (85–95% yield), and the gp41-8 intein $(7.7 \pm 1.5) \times 10^{-5}$ s^{-1}, $t_{1/2}$ ~150 min (40–50% yield). The precise C-cleavage was confirmed by Edman sequencing of the liberated 5aa^C-Trx (data not shown).

Table 4 also shows the effect of temperature on the C-cleavage reaction. Lower rates were associated with lower temperatures, but the yields were only slightly compromised. For example, at 0 °C, the gp41-1, NrdJ-1, and IMPDH-1 inteins showed yields as high as at 37 °C (>80%), meaning that cleavage efficiencies are robust over a wide temperature range. On the other hand, the gp41-8 intein generally showed lower rates with lower yields (<50%). In addition, analysis of the pH influence indicated that pH values above 7 resulted in a decrease of C-cleavage efficiency (data not shown). The best C-cleavage yields and constant rates under the evaluated conditions were obtained at 37 °C and pH 7 (see also Fig. 4*E*).

In summary, the four new inteins could be used for C-terminal cleavage reactions by introducing a C1A mutation, although the rates and in some cases the yields were markedly reduced when compared with the *trans*-splicing reaction. This finding is in contrast to the naturally split inteins *Npu* DnaE and *Ssp* DnaE whose C-terminal cleavage reaction is completely blocked in the context of a C1A mutation (12, 26).

C-cleavage without Any Extra Amino Acids and Effect of Native Extein Sequence—For practical applications, it is often highly desirable to have a clean cleavage without any additional amino acid attached to the product of interest. We generated four clones without the 5 residues corresponding to the native

TABLE 4
Apparent first order rate constants and yields of the C-terminal cleavage with five native Ext^C amino acids (+5aa^C)
ND = not determined; k = constant rate; Y = yield. Observed rates and yields were calculated as described under "Experimental Procedures." The time-courses for the C-terminal cleavage reaction were monitored over 24 h and commenced upon the addition of the N-domain to the C-domain (each construct at 5 μM).

| | gp41-1 (11 + 6) | | gp41-8 (12 + 7) | | NrdJ-1 (13 + 8) | | IMPDH-1 (14 + 9) | |
Temperature	k	Y	k	Y	k	Y	k (s^{-1})	Y
°C	*s^{-1}*	*%*	*s^{-1}*	*%*	*s^{-1}*	*%*	*s^{-1}*	*%*
0	$(5.7 \pm 2.8) \times 10^{-5}$	80–90	$(1.0 \pm \text{ND}) \times 10^{-5}$	30–40	$(5.0 \pm \text{ND}) \times 10^{-5}$	85–95	$(2.0 \pm \text{ND}) \times 10^{-5}$	85–95
25	$(5.4 \pm 0.5) \times 10^{-4}$	85–95	$(6.0 \pm 2.0) \times 10^{-5}$	30–40	$(1.7 \pm 0.8) \times 10^{-4}$	90–95	$(3.1 \pm 1.1) \times 10^{-4}$	85–95
37	$(2.2 \pm 0.7) \times 10^{-3}$	85–95	$(7.7 \pm 1.5) \times 10^{-5}$	40–50	$(2.4 \pm 0.1) \times 10^{-4}$	85–95	$(7.7 \pm 1.6) \times 10^{-4}$	85–95

C-extein sequence (proteins 15–18) and tested for C-cleavage activity when the purified proteins were mixed with the complementary Int^N C1A fusions (proteins 11–14). In Fig. 4, the *right panels* show these cleavage reactions for each intein. Fig. 4*A*, *right panel*, the complete time course of the C-cleavage mediated by the gp41-1 intein is shown. The Int^C precursor (*band 15*) is almost completely cleaved, liberating the Trx (without any extra amino acid) and Int^C fragment. For all the inteins, the predicted N terminus of the Trx cleavage product was confirmed by Edman sequencing (data not shown).

The optimal temperature for each intein was determined. As can be seen in Table 5, in general, the four split inteins showed high thermostability over a wide range of temperatures (0–55 °C), with the highest C-cleavage activities between 37 and 50 °C.

The deletion of the 5 native C-extein residues affected the rates and yields of the four split inteins at varying degrees (Tables 4 and 5, Fig. 4*E*). At 37 °C, the gp41-1 intein had a 10-fold rate reduction but maintained its excellent yields (85–95%). The gp41-8 intein was again associated with lower yields (20–30%), and it had a 1.5-fold rate reduction, whereas the NrdJ-1 intein had a 5-fold rate reduction and a drop to 60–70% yield. The IMPDH-1 intein was severely affected, having a 38.5-fold rate reduction and a drop to 65–75% yield. Similar to the *trans*-splicing results (Table 2), the gp41-1 and IMPDH-1 inteins were active even at 60 °C, whereas the gp41-8 and NrdJ-1 intein were inactive at temperatures higher than 50 °C.

In conclusion, under all evaluated conditions, the gp41-1 intein possessed the highest rates and yields among the four inteins. It can be considered as the most robust split intein for C-cleavage without adding extra native amino acids.

DISCUSSION

Split inteins hold a tremendous potential for a wide range of protein engineering approaches (17, 19, 30). Recently, novel biotechnological applications for split inteins were described. They represent a combination of affinity binding and cleavage for protein purification (31), the generation of antibody toxin conjugates (32), or widespread use as tools in drug discovery (33). In general, intein-tagged fusion protein systems enhance the efficiency of protein affinity purification processes by way of requiring only a single chromatographic step (34). Thus, there is a great need for novel split inteins with beneficial biochemical properties.

We characterized for the first time the functionality, in terms of protein splicing and self-cleavage activities, of four novel hypothetical split inteins, gp41-1, gp41-8, NrdJ-1, and

IMPDH-1, previously predicted from metagenomic sequences (27). All pairs of split intein protein fragments were assessed as model fusion proteins under identical extein context to allow their comparison with the previously reported extremely efficient and fast-splicing split intein *Npu* DnaE (26). Splicing rates of the gp41-1, gp41-8, NrdJ-1, and IMPDH-1 inteins were 10-, 3-, 7-, and 6-fold higher, respectively, than those reported for the *Npu* DnaE at 37 °C, and all of them showed excellent (90–95%) splicing yields (Table 2) with no apparent side-product formation. It is also noteworthy that both precursor halves were nearly quantitatively consumed in the reactions. Our results for the *Npu* DnaE *trans*-splicing rates were similar to previous studies (26), confirming that the observed rate differences cannot be accounted for by different laboratory conditions of any kind.

To further characterize the naturally split inteins, we tested their splicing activity over a wide range of temperatures (0–60 °C). Interestingly, we found that the optimal temperature for the gp41-1 split intein was around 45 °C, whereas for gp41-8, NrdJ-1, and IMPDH-1 inteins, it was 37 °C (Fig. 3). The functional thermostability of these split inteins might be explained by the environmental conditions of the sampling site. The genes corresponding to all four novel split inteins were identified from an environmental microbial source GOS metagenomic data sampled in Lake Gatun (Panama), near the tropical Caribbean Sea, with water temperatures of 28.5 °C (35). The gp41-1 intein was found to be most robust and gave excellent yields at a wide range of conditions, including different alkali pHs, high temperatures such as 60 °C, and the presence of chaotropic agents.

We observed an absence of promiscuity in different exogenous intein fragment combinations, *i.e.* only the complementary pairs of inteins fragments were active. Such orthogonality would hinder the functional complementation of different split inteins inside the same host cell and therefore would not exclude the possibility that some or all of these inteins are from the same microbial origin. However, given their isolation from large-scale metagenomic sequencing efforts, it appears more likely that the four inteins are derived from different organisms. For some biotechnological applications, this orthogonality could be highly valuable as it allows their simultaneous use, for example, for three-piece ligations (29, 36). In contrast, most exogenous combinations of cyanobacterial DnaE intein alleles exhibit efficient cross-splicing activity (7, 8) and are therefore less suited for such purposes.

Another example of a potential application for the new split inteins is protein labeling using the Cys tag approach, in which
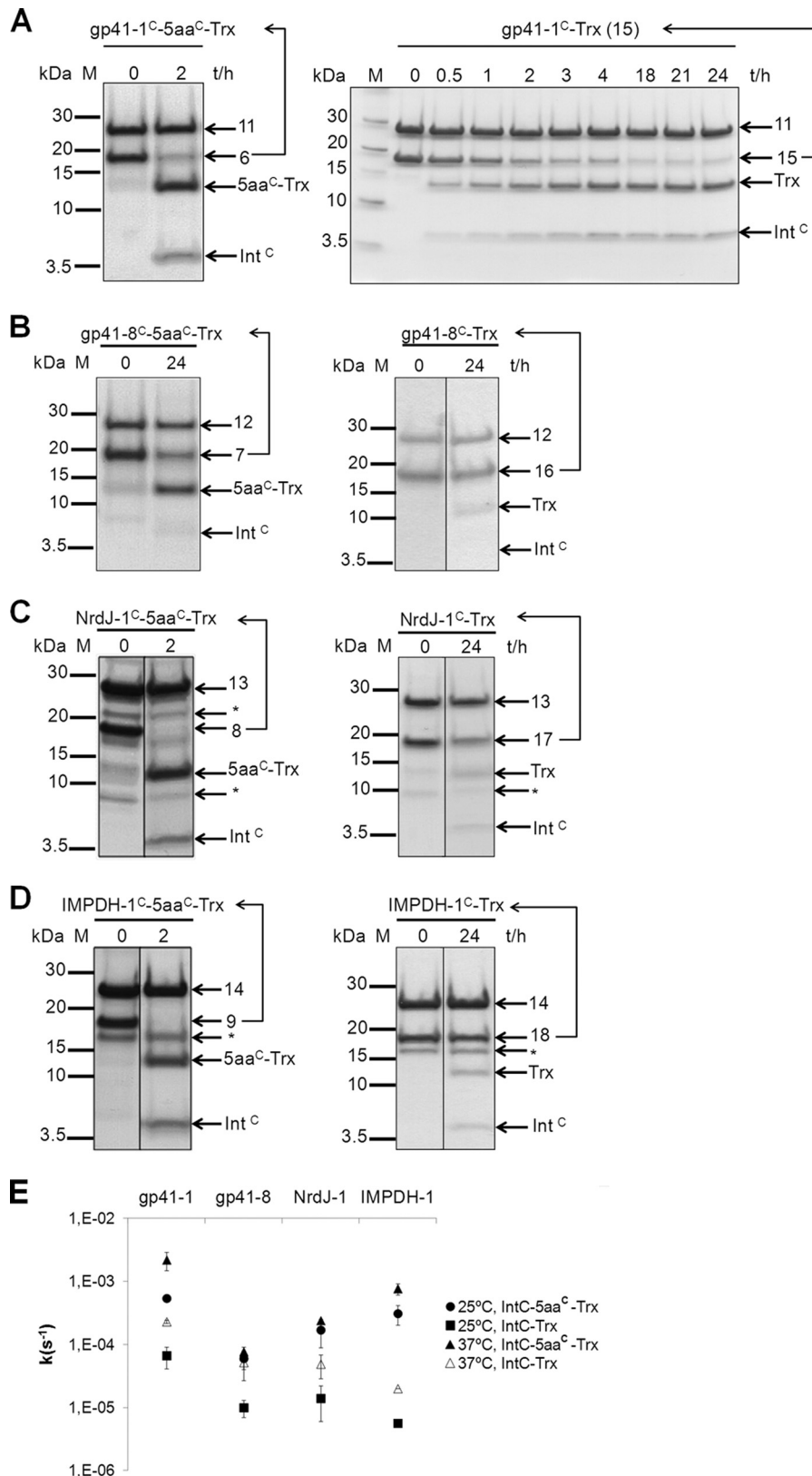
FIGURE 4. **C-cleavage with and without native extein amino acids.** For these reactions at 37 °C, a C1A mutation was introduced into each Int[N] fusion protein. 5 native C-extein amino acids (5aa[C]) were either included (*left panels*) or deleted (*right panels*). *A–D*, shown are Coomassie Brilliant Blue-stained SDS-gels. The C-terminal precursor proteins (indicated by *numbers*) were either completely or partially converted into the Extein[C] (indicated as Trx in all cases) and the Intein[C] fragment. The N-terminal precursor protein remained unchanged during the reaction. Proteins bands marked with an *asterisk* denote protein impurities. *M* denotes molecular marker. *E*, C-cleavage constant rates of the four inteins determined at 25 and 37 °C. *Error bars* indicate S.D.

TABLE 5

Apparent first order rate constants and yields of C-terminal cleavage without any native Ext^C amino acids

NA = Not Active; ND = not determined; k = constant rate; Y = yield. Observed rates and yields were calculated as described under "Experimental Procedures." The time-courses for the C-terminal cleavage reaction were monitored over 24 h and commenced upon the addition of the N-domain to the C-domain (each construct at 5 μM).

	gp41-1 (11 + 15)		gp41-8 (12 + 16)		NrdJ-1 (13 + 17)		IMPDH-1 (14 + 18)	
Temperature	k	Y	k	Y	k	Y	k	Y
$^\circ C$	s^{-1}	%	s^{-1}	%	s^{-1}	%	s^{-1}	%
25	$(6.7 \pm 2.5) \times 10^{-5}$	85–95	$(1.0 \pm ND) \times 10^{-5}$	10–15	$(1.4 \pm 0.8) \times 10^{-5}$	30–40	$(5.7 \pm 0.2) \times 10^{-6}$	15–25
37	$(2.3 \pm 0.1) \times 10^{-4}$	90–95	$(5.2 \pm 2.5) \times 10^{-5}$	20–30	$(4.9 \pm 2.0) \times 10^{-5}$	60–70	$(2.0 \pm 0.0) \times 10^{-5}$	65–75
45	$(4.4 \pm 1.2) \times 10^{-4}$	90–95	$(6.7 \pm 1.5) \times 10^{-5}$	20–30	$(3.1 \pm 0.2) \times 10^{-5}$	70–80	$(2.8 \pm 0.1) \times 10^{-5}$	70–80
50	$(4.3 \pm 0.5) \times 10^{-4}$	85–95	$(6.1 \pm 0.1) \times 10^{-5}$	20–30	$(6.0 \pm ND) \times 10^{-6}$	10–15	$(3.5 \pm 0.7) \times 10^{-5}$	80–90
55	$(1.9 \pm 1.0) \times 10^{-4}$	65–85	NA		NA		$(8.0 \pm ND) \times 10^{-6}$	60–70
60	$(6.0 \pm ND) \times 10^{-5}$	40–50	NA		NA		$(8.3 \pm ND) \times 10^{-6}$	<10

.

a prelabeled cysteine is added to the protein of interest by protein *trans*-splicing (37, 38). Because all four inteins characterized here operate with a Ser+1 instead of a Cys+1 residue and are also free of Cys residues in the remainder of their Int^C sequences (except for a Cys-2 in the gp41-8 Int^C fragment), they would be ideally suited for this technology without the need for further mutagenesis.

Split inteins, like regular *cis*-inteins, can be genetically engineered to allow controllable *in vitro* N- or C-cleavage (21), which is useful for protein purification purposes, for example. A specific advantage of split inteins in this respect is that uncontrolled *in vivo* cleavage is circumvented, which may have caused poor yields of the target protein (39). We generated a C1A mutant for each of the four novel split inteins to test whether these proteins could be used for C-terminal cleavage applications. Previously reported naturally split inteins display very slow or no self-cleavage activity at all when mutated at the same position (26, 40). *Ssp* DnaE is the only reported natural split intein rendering a C-terminal cleavage rate of 1.9×10^{-4} s^{-1}, but only after N-terminal cleavage has occurred. This can be induced by thiolysis, thus indicating a clear coordination of steps (12). Intriguingly, although the unprecedented high rate constants and efficiency of the four new inteins in the *trans*-splicing reaction might have suggested a highly optimized active site that is likewise resistant to aberrant side reactions induced by single amino acid mutations, we find that this is not the case. In contrast to the split DnaE alleles, our experimental data indicate that the naturally split inteins gp41-1$^{N(C1A)}$, gp41-8$^{N(C1A)}$, NrdJ-1$^{N(C1A)}$, and IMPDH-1$^{N(C1A)}$ exhibited C-terminal cleavage, although to different extents. Under conditions at pH 7 and 37 °C, this reaction was about 60-, 580-, 410-, and 110-fold slower than the *trans*-splicing reaction of the corresponding split inteins, respectively. Among the four inteins, gp41-1 displayed the highest C-cleavage reaction rate (Table 4) with a nearly complete conversion (85–90% yield). The NrdJ-1 and IMPDH-1 inteins exhibit similar constant rates and yields of C-terminal self-cleavage under the same conditions, whereas the gp41-8 intein shows the slowest rate and lowest yield of C-cleavage (near 50%).

It is fairly well known that modifications of native extein residues directly flanking the intein sequence can lead to inefficient *trans*-splicing or self-cleavage reactions (41). However, additional residues on the target protein are undesirable for some recombinant protein production applications, therefore making their removal essential. The current study shows that the four described new split inteins retain their C-cleavage activity after deletion of their five C-terminal native extein residues, although at 37 °C, the C-terminal cleavage rates of the gp41-1, gp41-8, NrdJ-1, and IMPDH-1 inteins were decreased by 10-, 1.5-, 5-, and 38.5-fold, respectively (Table 5 and Fig. 4E). Removal of the native flanking residues also negatively affects, to differing degrees, the C-cleavage reaction yields of all split inteins except for the gp41-1 intein, which was still capable of driving the cleavage reaction to almost 90% precursor consumption.

The observed extraordinarily high constant rates of the novel inteins are intriguing and could not have been predicted before this study. The obvious prerequisites for this behavior are a highly optimized recognition and folding of the split intein fragments. For the association of natively split intein fragments, complementary local charge distributions appear to be an important parameter (7, 27, 29, 42, 43). The fragments of the *Ssp* DnaE inteins are positively and negatively charged and form a complex in solution with a rate approaching the diffusion limit (29), whereas the rate-determining step appears to be an intramolecular rearrangement within the complex. Thus, although it was previously postulated that the gp41-1, gp41-8, NrdJ-1 ,and IMPDH-1 inteins have 10, 7, 7, and 3 interfragment salt-bridges, respectively, in the predicted interaction interface between the two split parts (27), it appears unlikely that the association could be faster than that observed for the *Ssp* DnaE intein or at least that this could account for an increased overall reaction rate. However, we propose that the high rates observed here are a consequence of the rapid folding into the active intein structure, the formation of highly optimized active sites at the N- and C-terminal junctions, and an optimal coupling of the individual steps during protein splicing.

We are currently unable to predict such a superior catalytic behavior from the primary sequence of the intein. Recent studies on the directed evolution of inteins indicated that a few amino acid substitutions can have a significant positive effect on their performance (30, 44–47). As previously described (27), the new split inteins are not closely related to other established split and full-length inteins. In fact, their protein sequences only share up to 40% similarity to other known cyanobacterial naturally split inteins (7, 35). Among them, the gp41-1, gp41-8, NrdJ-1, and IMPDH-1 inteins also share low identity (40–50%), denoting relevant sequence variations in their intein halves.

ASBMB

The sequences and gene organization of the reported split inteins are closely related to those found in phages and viruses, suggesting their viral origin (27). It is known that inteins and split inteins contain six conserved protein splicing motifs of the HINT (Hog/Intein) family: N1 to N4 present in the N-terminal fragment of the split intein and C1 and C2 present in the C-terminal split intein fragment (5–7, 48). Of the six conserved motifs, N1, N3, C1, and C2 are the most conserved ones among the naturally split inteins and have been described as block A, B, F and G, respectively. All these sequence motifs can be identified in the gp41-1, gp41-8, NrdJ, and IMPDH-1 inteins. The only apparent major difference to the *Ssp* and *Npu* DnaE inteins is a histidine as the penultimate amino acid in the G box, which in the well characterized natively split cyanobacterial inteins is a serine (data not shown). However, in most other inteins, a His residue is also found at this position, which is postulated to assist in cyclization of the ultimate Asn residue to effect cleavage at the C-terminal splice junction (49). Thus, the new inteins characterized here exhibit canonical key catalytic residues and do not represent unusual features in their primary sequences.

Altogether these results show that the gp41-1 intein is superior on all accounts to those investigated in this study. It is very tolerant to a broad range of reaction conditions and tolerates changes in its splicing junction region for C-terminal cleavage reactions. Most importantly, when compared with all other split inteins reported so far, from natural, artificial, or engineered origin, the four inteins characterized in this study show the highest reaction rates and efficiencies in the *trans*-splicing reaction. In particular, the gp41-1 intein increases the limits of the fastest split intein ever described and holds great potential as a protein ligation tool or in self-cleavage processes for protein purification.

REFERENCES

1. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J., and Belfort, M. (1994) Protein splicing elements: inteins and exteins: a definition of terms and recommended nomenclature. *Nucleic Acids Res.* **22,** 1125–1127
2. Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1990) Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H^+-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265,** 6726–6733
3. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kDa subunit of the vacuolar H^+-adenosine triphosphatase. *Science* **250,** 651–657
4. Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W. R., Jr., and Perler, F. B. (1997) The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. *J. Bacteriol.* **179,** 6378–6382
5. Perler, F. B. (2002) InBase: the Intein Database. *Nucleic Acids Res.* **30,** 383–384
6. Pietrokovski, S. (1998) Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci.* **7,** 64–71
7. Dassa, B., Amitai, G., Caspi, J., Schueler-Furman, O., and Pietrokovski, S. (2007) *trans*-Protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. *Biochemistry* **46,** 322–330
8. Iwai, H., Züger, S., Jin, J., and Tam, P. H. (2006) Highly efficient protein *trans*-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.* **580,** 1853–1858
9. Wu, H., Hu, Z., and Liu, X. Q. (1998) Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 9226–9231
10. Wei, X. Y., Sakr, S., Li, J. H., Wang, L., Chen, W. L., and Zhang, C. C. (2006) Expression of split *dnaE* genes and *trans*-splicing of DnaE intein in the developmental cyanobacterium *Anabaena* sp. PCC 7120. *Res. Microbiol.* **157,** 227–234
11. Choi, J. J., Nam, K. H., Min, B., Kim, S. J., Söll, D., and Kwon, S. T. (2006) Protein *trans*-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*. *J. Mol. Biol.* **356,** 1093–1106
12. Martin, D. D., Xu, M. Q., and Evans, T. C., Jr. (2001) Characterization of a naturally occurring *trans*-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry* **40,** 1393–1402
13. Sun, W., Yang, J., and Liu, X. . (2004) Synthetic two-piece and three-piece split inteins for protein *trans*-splicing. *J. Biol. Chem.* **279,** 35281–35286
14. Ludwig, C., Pfeiff, M., Linne, U., and Mootz, H. D. (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
15. Appleby, J. H., Zhou, K., Volkmann, G., and Liu, X. Q. (2009) Novel split intein for *trans*-splicing synthetic peptide onto C terminus of protein. *J. Biol. Chem.* **284,** 6194–6199
16. Tavassoli, A., and Benkovic, S. J. (2007) Split intein-mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. *Nat. Protoc.* **2,** 1126–1133
17. Muralidharan, V., and Muir, T. W. (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Methods* **3,** 429–438
18. Volkmann, G., and Iwaï, H. (2010) Protein *trans*-splicing and its use in structural biology: opportunities and limitations. *Mol. Biosyst.* **6,** 2110–2121
19. Mootz, H. D. (2009) Split inteins as versatile tools for protein semisynthesis. *Chembiochem.* **10,** 2579–2589
20. Wu, W. Y., Mee, C., Califano, F., Banki, R., and Wood, D. W. (2006) Recombinant protein purification by self-cleaving aggregation tag. *Nat. Protoc.* **1,** 2257–2262
21. Mathys, S., Evans, T. C., Chute, I. C., Wu, H., Chong, S., Benner, J., Liu, X. Q., and Xu, M. Q. (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,** 1–13
22. Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H., and Xu, M. Q. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192,** 271–281
23. Tan, G., Chen, M., Foote, C., and Tan, C. (2009) Temperature-sensitive mutations made easy: generating conditional mutations by using temperature-sensitive inteins that function within different temperature ranges. *Genetics* **183,** 13–22
24. Chong, S., Montello, G. E., Zhang, A., Cantor, E. J., Liao, W., Xu, M. Q., and 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
25. Volkmann, G., Sun, W., and Liu, X. Q. (2009) Controllable protein cleavages through intein fragment complementation. *Protein Sci.* **18,** 2393–2402
26. Zettler, J., Schütz, V., and Mootz, H. D. (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction. *FEBS Lett.* **583,** 909–914
27. Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O., and Pietrokovski, S. (2009) Fractured genes: a novel genomic arrangement involving

new split inteins and a new homing endonuclease family. *Nucleic Acids Res.* **37,** 2560–2573

28. Tori, K., Dassa, B., Johnson, M. A., Southworth, M. W., Brace, L. E., Ishino, Y., Pietrokovski, S., and Perler, F. B. (2010) Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *J. Biol. Chem.* **285,** 2515–2526

29. Shi, J., and Muir, T. W. (2005) Development of a tandem protein *trans*-splicing system based on native and engineered split inteins. *J. Am. Chem. Soc.* **127,** 6198–6206

30. Appleby-Tagoe, J. H., Thiel, I. V., Wang, Y., Wang, Y., Mootz, H. D., and Liu, X. Q. (2011) Highly efficient and more general *cis*- and *trans*-splicing inteins through sequential directed evolution. *J. Biol. Chem.* **286,** 34440–34447

31. Lu, W., Sun, Z., Tang, Y., Chen, J., Tang, F., Zhang, J., and Liu, J. N. (2011) Split intein facilitated tag affinity purification for recombinant proteins with controllable tag removal by inducible auto-cleavage. *J. Chromatogr. A* **1218,** 2553–2560

32. Möhlmann, S., Bringmann, P., Greven, S., and Harrenga, A. (2011) Site-specific modification of ED-B-targeting antibody using intein fusion technology. *BMC Biotechnol.* **11,** 76

33. Cheriyan, M., and Perler, F. B. (2009) Protein splicing: A versatile tool for drug discovery. *Adv. Drug Deliv. Rev.* **61,** 899–907

34. Fong, B. A., Wu, W. Y., and Wood, D. W. (2010) The potential role of self-cleaving purification tags in commercial-scale processes. *Trends Biotechnol.* **28,** 272–279

35. Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y. H., Falcón, L. I., Souza, V., Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Nealson, K., Friedman, R., Frazier, M., and Venter, J. C. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5,** e77

36. Busche, A. E., Aranko, A. S., Talebzadeh-Farooji, M., Bernhard, F., Dötsch, V., and Iwaï, H. (2009) Segmental isotopic labeling of a central domain in a multidomain protein by protein *trans*-splicing using only one robust DnaE intein. *Angew. Chem. Int. Ed. Engl.* **48,** 6128–6131

37. Kurpiers, T., and Mootz, H. D. (2007) Regioselective cysteine bioconjugation by appending a labeled cysteine tag to a protein by using protein splicing in *trans*. *Angew. Chem. Int. Ed. Engl.* **46,** 5234–5237

38. Kurpiers, T., and Mootz, H. D. (2008) Site-specific chemical modification of proteins with a prelabeled cysteine tag using the artificially split Mxe GyrA intein. *Chembiochem.* **9,** 2317–2325

39. Li, Y. (2011) Self-cleaving fusion tags for recombinant protein production. *Biotechnol. Lett.* **33,** 869–881

40. Nichols, N. M., and Evans, T. C., Jr. (2004) Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split Ssp DnaE intein. *Biochemistry* **43,** 10265–10276

41. Amitai, G., Callahan, B. P., Stanger, M. J., Belfort, G., and Belfort, M. (2009) Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 11005–11010

42. Brenzel, S., Kurpiers, T., and Mootz, H. D. (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

43. Shah, N. H., Vila-Perelló, M., and Muir, T. W. (2011) Kinetic control of one-pot *trans*-splicing reactions by using a wild-type and designed split intein. *Angew. Chem. Int. Ed. Engl.* **50,** 6511–6515

44. Wood, D. W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999) A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotechnol.* **17,** 889–892

45. Lockless, S. W., and Muir, T. W. (2009) Traceless protein splicing utilizing evolved split inteins. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 10999–11004

46. Hiraga, K., Soga, I., Dansereau, J. T., Pereira, B., Derbyshire, V., Du, Z., Wang, C., Van Roey, P., Belfort, G., and Belfort, M. (2009) Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. *J. Mol. Biol.* **393,** 1106–1117

47. Peck, S. H., Chen, I., and Liu, D. R. (2011) Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells. *Chem. Biol.* **18,** 619–630

48. Elleuche, S., and Pöggeler, S. (2010) Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl. Microbiol. Biotechnol.* **87,** 479–489

49. Chen, L., Benner, J., and Perler, F. B. (2000) Protein splicing in the absence of an intein penultimate histidine. *J. Biol. Chem.* **275,** 20431–20435