BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

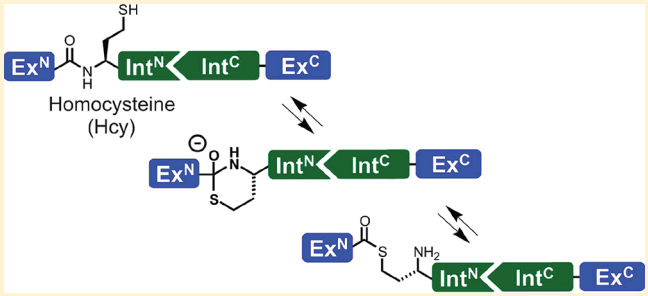
pubs.acs.org/biochemistry

Probing Intein-Catalyzed Thioester Formation by Unnatural Amino Acid Substitutions in the Active Site

Dirk Schwarzer,[†] Christina Ludwig,[‡] Ilka V. Thiel, and Henning D. Mootz*

Institute of Biochemistry, University of Muenster, Wilhelm-Klemm-Strasse 2, 48149 Münster, Germany

ABSTRACT: Inteins are single-turnover catalysts that splice themselves out of a precursor polypeptide chain. For most inteins, the first step of protein splicing is the formation of a thioester through an N−S acyl shift at the upstream splice junction. However, the mechanism by which this reaction is achieved and the impact of mutations in and close to the active site remain unclear on the atomic level. To investigate these questions, we have further explored a split variant of the *Ssp* DnaB intein by introducing substitutions with unnatural amino acids within the short synthetic N-terminal fragment. A previously reported collapse of the oxythiazolidine anion intermediate into a thiazoline ring was found to be specifically dependent on the methyl side chain of the flanking Ala(−1). The stereoisomer D-Ala and the constitutional isomers *β*-Ala and sarcosine did not lead to this side reaction but rather supported splicing. Substitution of the catalytic Cys1 with homocysteine strongly inhibited protein splicing; however, thioester formation was not impaired. These results argue against the requirement of a base to deprotonate the catalytic thiol group prior to the N−S acyl shift, because it should be misaligned for optimal proton abstraction. A previously described mutant intein evolved for more general splicing in different sequence contexts could even rather efficiently splice with this homocysteine. Our findings show the large impact of some subtle structural changes on the protein splicing pathway, but also the remarkable tolerance toward other changes. Such insights will also be important for the biotechnological exploitation of inteins.



Inteins are internal protein domains that excise themselves out of a precursor protein under concomitant linkage of the N- and C-terminal polypeptide sequences, termed exteins. This process is called protein splicing.[1−4] The intein is a self-contained entity as it requires no additional factors or energy sources. Rather, protein splicing proceeds in an autocatalytic fashion and commences as soon as the intein domain is correctly folded. The same is true for split inteins, in which N- and C-terminal fragments of the intein domain (Int^N and Int^C, respectively) are localized on separate polypeptide chains and associate with each other to reconstitute the active intein prior to protein *trans*-splicing. These properties as well as the tolerance of *cis*- and *trans*-splicing inteins for many heterologous sequence contexts allowed for many biotechnological applications of inteins, including protein purification, protein semisynthesis, segmental isotopic labeling, and polypeptide cyclization.[5−8]

The protein splicing pathway represents a remarkable sequence of rearrangements affecting the peptide backbone at the junctions between intein and both exteins. These reactions must be highly coordinated to prevent side reactions of the reactive intermediates. Inteins show generally a low degree of sequence homology but share several block motifs with conserved amino acids that are crucial for the reaction mechanism.[1,9,10] Briefly, in the first step, the side chain of the first residue of the intein in the block A motif, which is either a Cys or a Ser, attacks the upstream carbonyl carbon atom of the scissile peptide bond to effect an N−S or N−O acyl shift through a thiazolidine or oxazolidine intermediate. The N-extein acyl chain of the resulting thioester or oxoester is then transferred onto the nucleophilic side chain of the first residue of the C-extein at the C-terminal intein-extein junction in the block G motif, which can be a Cys, Ser, or Thr. In the resulting branched intermediate, the side chain amide of the neighboring Asn, the last residue of the intein, attacks the carbonyl group of its own backbone amide, cyclizes to the succinimide, and facilitates cleavage of the C-terminal intein junction. Finally, a spontaneous S−N or O−N acyl shift rearranges the thioester or oxoester bond between the two exteins to the thermodynamically favored peptide bond. While this represents the canonical pathway of protein splicing, there is also a subset of inteins that have an amino acid other than Cys or Ser at their first position and thus cannot undergo an acyl shift reaction at the upstream scissile bond. In these cases, attack of the downstream nucleophile is believed to occur directly on the peptide bond.[11,12]

The formation of the active ester intermediate at the upstream splice junction without the consumption of any energy sources can be regarded as the most intriguing reaction in protein splicing. Although different inteins seem to have evolved slightly different mechanisms to overcome this energy barrier, a distortion or polarization of the scissile peptide bond appears to be a common strategy. Structural studies suggest that the strictly conserved

Received: September 22, 2011
Revised: December 7, 2011
Published: December 19, 2011

   dx.doi.org/10.1021/bi2014823 | *Biochemistry* 2012, 51, 233−242

His residue in the block B motif polarizes the peptide bond by hydrogen bonding. Mutation of this residue was shown to block the N−S acyl shift in many inteins.[1] Interestingly, in contrast to cysteine or serine proteases, no general base is obvious from structural investigations that might prime the Cys1 or Ser1 for nucleophilic attack through deprotonation, although a recent study indicated that the block B histidine might also play this role.[13] Structural studies using protein crystallography or nuclear magnetic resonance (NMR) also suggested a ground-state destabilization of the scissile peptide bond.[14,15] However, because of the self-processing nature of inteins, all structural information available was gained either from spliced inteins or from precursor proteins that were mutated at critical positions. Thus, the mechanism of the N−S or N−O shift still remains unclear at the atomic level.

We have recently reported an artificially split semisynthetic intein based on the *Ssp* DnaB mini-intein that consists of an 11-amino acid N-terminal fragment (Int^N) and a 143-amino acid C-terminal fragment (Int^C). Because of the short length of the Int^N piece, Ex^N-Int^N peptides with a short N-extein sequence can be easily accessed by solid-phase peptide synthesis. Upon mixture of the synthetic Ex^N-Int^N peptide and the recombinant Int^C-Ex^C polypeptide under native buffer conditions, the active intein is spontaneously formed.[16,17] This intein is a useful tool for preparing N-terminally modified semisynthetic proteins by protein *trans*-splicing. Furthermore, it also represents an attractive system for mechanistic studies because the upstream splice junction, comprising the catalytic Cys1 and the flanking N-extein residue Gly(−1), is localized on the synthetic peptide.[17,18] As such, we could detect the formation of a thiazoline ring at the upstream splice junction in the context of a mutant intein and a Gly(−1)Ala substitution.[17] This aberrant side product underscored the generally accepted notion that the N−S acyl shift proceeds via the oxythiazolidine anion intermediate.[19]

In this work, we have further explored the semisynthetic *Ssp* DnaB intein in mechanistic studies of the initial N−S acyl shift. Using $H_2{}^{18}O$ exchange, we investigated the formal possibility that the thiazoline might constitute a regular intermediate in the splicing pathway. Replacing the Gly(−1) residue with unnatural amino acids that are constitutional, functional, and stereoisomers of Ala provided further insight into the requirements at this critical position. Finally, substituting Cys1 with D-Cys or homocysteine (Hcy) allowed us to probe the influence of slight structural and electronic changes at this catalytically essential amino acid. This is the first investigation of the upstream splicing junction using unnatural amino acids. Our findings reveal that subtle changes at the critical positions either can be tolerated or can abrogate the protein splicing pathway at different steps. Implications for the catalytic strategy of the N−S acyl shift are discussed.

MATERIALS AND METHODS

General Techniques and Materials. Unless otherwise specified, standard protocols were used. Reagents were purchased from Novabiochem, Sigma-Aldrich, Arcos, or Roth. All reactions and assays were performed in at least duplicate.

Generation of Protein Constructs and Peptide Synthesis. The expression plasmids, protein expression conditions, and protein purification procedures for constructs **1**−**3** and construct **1a** were described previously.[16,17,20] Peptides were synthesized using standard Fmoc-based solid-phase chemistry on a Syro XP synthesizer (Multisyntech). Wang resin was used as solid support, and amino acid side chains were protected as follows: Asp(tBu), Glu(tBu), Lys(Boc), and Ser(tBu). Standard coupling reactions were performed by activating 5 equiv of the Fmoc-protected amino acid derivative with 4.8 equiv of 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) in the presence of 6 equiv of *N*,*N*-diisopropylethylamine (DIPEA) in DMF as a base. Removal of the Fmoc group was conducted with 20% piperidine in DMF. The L-Cys (in pep**1**−pep**7**) as well as the D-Cys building block (in pep**9**) were coupled manually as pentafluorophenyl esters (Novabiochem) without addition of base, to prevent even small amounts of cysteine racemization. The Hcy moiety in pep**8** was introduced using HBTU-activated *N*-Fmoc-*S*-trityl-L-homocysteine (Bachem) and 6 equiv of *N*-methylmorpholine (NMM). HBTU-activated 5,6-carboxyfluorescein was coupled twice manually for at least 8 h in the presence of 6 equiv of NMM. The following unnatural amino acid building blocks were also introduced by the standard coupling protocol using HBTU and NMM: Fmoc-*β*-Ala-OH (Novabiochem), Fmoc-Sar-OH (Novabiochem), and Fmoc-D-Ala-OH (Novabiochem). Peptides were cleaved off the resin with complete removal of the side chain protecting groups by a cleavage cocktail that contained TFA, H_2O, thioanisole, 1,2-ethanedithiol, and phenol (85:5:5:2.5:2.5) for 4 h. Cleaved products were precipitated in cold diethyl ether, centrifuged and washed with diethyl ether, dissolved in H_2O, and lyophilized. Crude peptides were purified via high-performance liquid chromatography (HPLC) on a preparative C18 column.

Assays for Protein Splicing and Cleavage. A detailed description of the reaction conditions and the densitometric analysis of the Coomassie-stained gels was previously published.[16,17] In brief, peptides and proteins were mixed at the indicated concentrations, typically between 20 and 68 μM, in assay buffer [50 mM Tris, 300 mM NaCl, 2 mM dithiothreitol (DTT), 1 mM ethylenediaminetetraacetic acid (EDTA), and 10% glycerine (pH 7.0)] and incubated at 25 °C for 24 h. Removed aliquots of the splicing reaction mixtures were quenched by addition of 4× SDS loading buffer (containing 8% SDS and 20% *β*-mercaptoethanol) and boiled before being applied to a sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) gel. Aliquots from the intein complex formation assays were quenched with sample buffer, but the protein samples were not boiled before being loaded onto the gel. Gels were stained with Coomassie brillant blue, and relative intensities of protein bands were densitometrically determined using Scion Image (http://www.scioncorp.com). For UV illumination, the SDS gels were photographed under UV light before being stained.

^{18}O Isotope Incorporation Assays. The ^{18}O buffer was produced by directly dissolving 50 mM Tris, 300 mM NaCl, and 2 mM tris(2-carboxyethyl)phosphine (TCEP) in ^{18}O isotopically marked water (Isotec), followed by adjustment of the pH to 7.0 with 37% HCl. As a control, a ^{16}O buffer was produced in exactly the same way. Highly concentrated stock solutions of peptides and proteins (in $H_2{}^{16}O$) were mixed with ^{18}O buffer, resulting in a final $H_2{}^{18}O$ content of 70% in the assay mixtures.

Matrix-Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-TOF MS) Analysis. MALDI-TOF MS measurements were performed on an Autoflex 2 instrument (Bruker Daltonics) in positive ion detection mode. To analyze peptide masses, the assay samples were mixed with a saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA)

in a ratio of 1:1 and directly subjected to the MS measurement. To study masses of larger proteins, the reaction mixtures were subjected to SDS−PAGE, and the desired Coomassie-stained protein bands were excised and trypsinized before being analyzed by MALDI-TOF MS using 0.2 μg of trypsin (Sigma-Aldrich) and incubated at 37 °C for 18 h.

Fluorescence Anisotropy Spectroscopy. Fluorescence anisotropy spectroscopy was conducted with a SafireII spectrofluorometer (Tecan). The fluoresceine-labeled peptides were incubated at a fixed concentration of 1 μM with various concentrations of construct **3** (0.05−30 μM). The increase in anisotropy was monitored over time, and the data were fit as described previously.[17,21]

Kinetic Investigation of the N-Terminal Hydrolysis Reaction. pep**8** (Cys1Hcy) or pep**1** (Cys1) at 100 μM were incubated with mutant construct **2** (55 μM) in assay buffer. At different time points (5 min, 50 min, 1.75 h, 4 h, 6.5 h, and 23 h), aliquots were removed and analyzed by analytical HPLC (Agilent 1100 series) under acidic conditions (buffer A consisting of H_2O and 0.1% TFA; buffer B consisting of acetonitrile and 0.1% TFA) with a Zorbrax 300SB C3 column (Agilent). All eluting signals were manually collected and further verified by MALDI-TOF MS. Integrated signal intensities of the hydrolysis products were plotted versus reaction time, and the data were fit to a second-order rate equation.

RESULTS

Strategy. We chose the semisynthetic split *Ssp* DnaB intein for mechanistic studies of the protein splicing reaction, because it offers several advantages. First, the beginning of the splice reaction can be controlled by mixing the Ex^N-Int^N and Int^C-Ex^C peptides or proteins. Second, because parts of the active site involving the N-terminal splicing junction are located on the short Int^N fragment, these residues are amenable to organic peptide synthesis and can be changed to building blocks beyond the genetically encoded amino acids. This allows one to introduce small structural changes with a high degree of precision. Furthermore, the Int^N fragment and modification or cleavage products thereof lend themselves to facile analysis by HPLC and MS because of their small size.

Four protein constructs were used in this study (for an overview, see Figure 1). Int^C-Trx (construct **1**) is the Int^C fragment (amino acids 12−143) fused to His-tagged thioredoxin (Trx). This construct is competent for protein *trans*-splicing when incubated with a suitable Ex^N-Int^N peptide.[16] Int^C-Trx_AA (construct **2**) differs from construct **1** in mutations of the two residues around the C-terminal splice junction, Asn154Ala and Ser(+1)Ala, which block all steps of the splicing pathway beyond the initial N−S acyl shift. It serves to investigate thioester formation.[17] Int^C-Trx_AAA (construct **3**) harbors an additional His73Ala substitution, which blocks thioester formation, but still allows the study of intein fragment association.[17] Finally, Int^{CM86}-Trx (construct **1a**) contains an evolved Int^C fragment that shows a higher splicing efficiency and a higher promiscuity toward flanking extein residues.[20] While the calculated masses of these constructs are 30.2−30.4 kDa, they migrated on SDS−PAGE gels at higher apparent molecular masses of ∼33−36 kDa in agreement with previous observations.[17,20]

Thiazoline Ring Formation Is Not a Step in the Regular Protein Splicing Pathway. In previous work, we identified the formation of a thiazoline ring arising from the

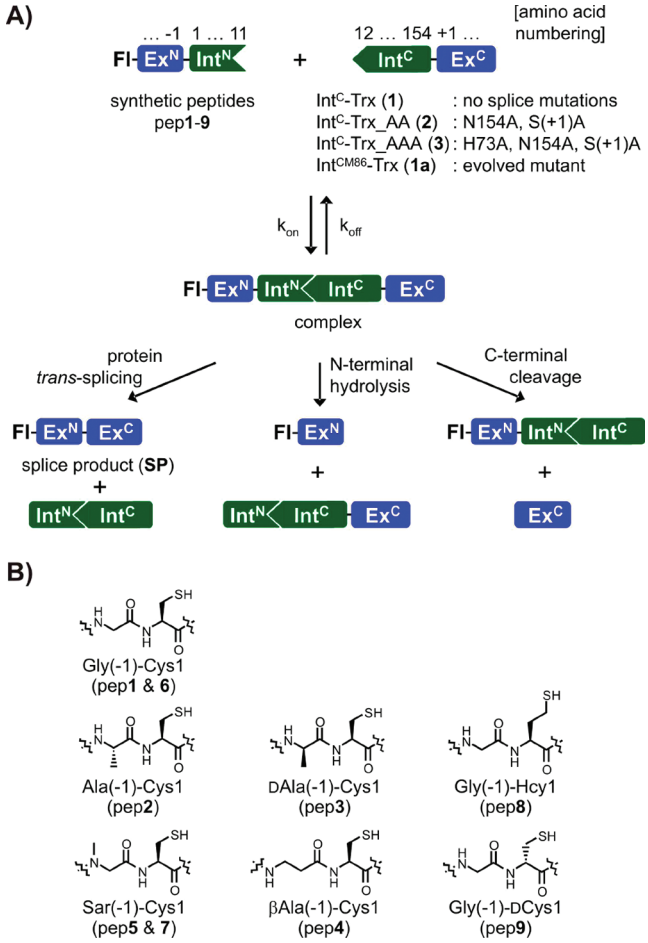


Figure 1. Strategy for studying the impact of changes at the N-terminal splice junction. (A) Shown are intein constructs Ex^N-Int^N and Int^C-Ex^C used in this study and their possible reaction pathways following initial formation of the intein complex. (B) Chemical structures of the amino acids flanking the scissile peptide bond at the Ex^N-Int^N junction in the peptides used in this study.

scissile peptide bond at the N-terminal splice junction.[17] It was observed when the Gly(−1)Ala mutation in the N-extein was combined with the Asn154Ala/Ser(+1)Ala substitutions in Int^C-Trx_AA (construct **2**), but not when the Int^C-Ex^C fragment was the wild-type sequence of Int^C-Trx (construct **1**). It therefore seemed that the unique combination of mutations led to a distortion of the active site that promoted aberrant water elimination of the oxythiazolidine anion intermediate to give the thiazoline. However, this observation did not rule out the possibility that the thiazoline is regularly occurring in the active site following the N−S acyl shift through reversible formation from the oxythiazolidine anion intermediate (Figure 2), and that only the kinetics or the equilibrium of the individual steps had so far precluded its isolation in the wild-type context.

To investigate this formal possibility, we used $H_2{}^{18}O$ to detect potential back-incorporation of water into the putative thiazoline intermediate. To establish such assays, we first tested the efficiency of ^{18}O incorporation by N-terminal hydrolysis caused by the attack of water on the thioester intermediate. Toward this end, an Ex^N-Int^N peptide with glycine at position −1 [pep**1** (Table 1)] was incubated with the Int^C-Trx_AA mutant (construct **2**) in buffer with either $H_2{}^{16}O$ or $H_2{}^{18}O$. The formation of the N-terminal cleavage products was analyzed by MALDI-TOF MS (see Figure 3A). As expected, the hydrolyzed

dx.doi.org/10.1021/bi2014823 | *Biochemistry* 2012, 51, 233−242

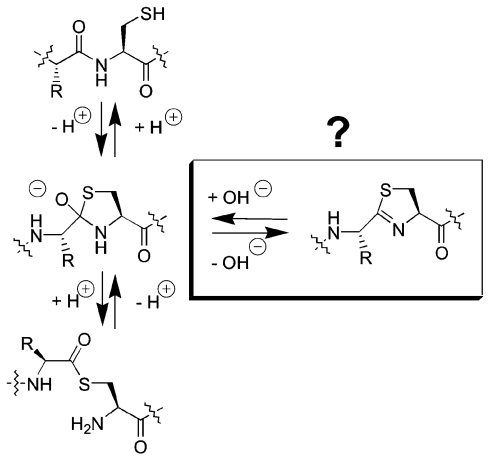


Figure 2. Mechanism of the N−S acyl shift. The peptide bond at the Ex^N-Int^N junction rearranges through the oxythiazolidine anion intermediate to the linear thioester. Following protonation, aberrant elimination of water can lead to the thiazoline (see the boxed reaction; here shown as hydroxide removal). Isotope labeling experiments with $H_2{}^{18}O$ were conducted to investigate the potential reversible formation of this product.

Ex^N was detected as Fl-KKESG-^{16}OH when the reaction was performed in $H_2{}^{16}O$, and mostly as Fl-KKESG-^{18}OH when the reaction was performed in $H_2{}^{18}O$. The approximate 3:7 ratio between the MS signals of the two products reflected the total $H_2{}^{18}O$ content in the buffer of 70%.

Using an Ex^N-Int^N peptide with alanine at position −1 (pep**2**) and Int^C-Trx (construct **1**), we then tested whether the thiazoline ring is potentially formed transiently and reversibly under these conditions (Figure 3B). This combination of intein fragments significantly impairs the protein *trans*-splicing reaction [see ref 17 and below (Table 2)]. In the case of reversible thiazoline formation, an increased level of incorporation of ^{18}O into pep**2** should be observed over time. However, only the mass of the unlabeled starting material was found under both reaction conditions (Figure 3B). This finding indicated that the significant loss of splicing activity is not due to a block of the splicing pathway caused by a (reversible) sink leading to the thiazoline.

Finally, we performed the splicing reaction between the Gly(−1) Ex^N-Int^N peptide pep**6** (which is functionally equivalent to pep**1**) and Int^C-Trx (construct **1**), which led to high yields of the splice product within 24 h. To analyze potential ^{18}O labeling at the carbonyl group of the newly formed Ex^N-Ex^C peptide bond, we conducted tryptic digestion and MALDI-TOF analysis of the splice product. However, in both reactions, only the mass of the unlabeled product could be detected (see Figure 3C), indicating that also under wild-type splice conditions no thiazoline ring is transiently formed.

Subtle Changes at Position −1 Can Have Large Effects on Protein Splicing and Thioester Formation. To rationalize the effect of the additional methyl group in the Gly(−1)Ala substitution, we consulted the crystal structure of the *Ssp* DnaB intein.[22] As illustrated in Figure 4A, the α-C atom of the native Gly(−1) is located only 3.3 Å from the π-N atom of the catalytic His73 residue. A model with the Gly(−1)Ala substitution indicated that the distance to the additional methyl group would be reduced to 2.3 Å, which is well below the van der Waals distance.[23] The steric clash could cause distortion of the active site significant enough to reduce the splicing activity and to cause thiazoline formation (see Figure 4B). We also noticed that the methyl side chain of a modeled D-Ala residue would not point toward the active site of the intein or collide with any other amino acids (see Figure 4C). We thus hypothesized that an extra methyl group at position −1 might be tolerated, however, not in the *S* configuration of L-Ala.

To test this assumption, we synthesized three peptides that also contained an additional methyl group yet should reduce the steric clash of residue −1 with His73. One peptide contained D-Ala in position −1 based on the consideration mentioned above [pep**3** (see Table 1)]. The other two peptides contained either β-Ala (pep**4**) or sarcosine (*N*-methylglycine, pep**5**) at position −1. In these latter two cases, the methyl group is formally inserted into the backbone or shifted to the upstream amide nitrogen, respectively. Peptides pep**3**−pep**5** were subjected to the *trans*-splicing reaction with construct **1**, and yields of the splice product were analyzed by SDS−PAGE and compared. Figure 5A shows that in each case the splice products Fl-KKESX-Trx-His_6 (X = Gly, Ala, D-Ala, β-Ala, or Sar, depending on the peptide) and the C-terminal cleavage byproduct Trx-His_6 were formed. The former could also be detected under UV illumination of the SDS−PAGE gel prior to Coomassie staining (Figure 5A, right panel). The amounts of formed splice products were quantified by densitometric analysis of the stained gel, and the amount obtained in the case of reference peptide pep**1** was set to 100%. In comparison, the reaction with the Ala-containing peptide pep**2** yielded only 7% splice product (see Table 2), which is consistent with our previous observations.[17] Interestingly, all three peptides with D-Ala, β-Ala, and Sar (pep**3**−pep**5**, respectively) gave rise to significantly higher yields of 26−38% (see Figure 5A and Table 2). Compared to pep**2**, they also resulted in better ratios of splice product relative to C-terminal cleavage product (see Figure 5A).

Table 1. Peptides Used in This Study

peptide	sequence[a]	mass (Da)	k_{on} (M^{-1} s^{-1})	K_d (μM)	calcd k_{off} ($\times 10^{-5}$ s^{-1})
pep**1**	Fl-KKESG-CISGDSLISLA	1964.9	16.8 ± 1.0	1.1 ± 0.2	1.8
pep**2**	Fl-KKES**A**-CISGDSLISLA	1978.9	22.0 ± 0.7	2.3 ± 0.1	5.1
pep**3**	Fl-KKES(D-**Ala**)-CISGDSLISLA	1978.9	10.3 ± 0.4	3.3 ± 0.1	3.4
pep**4**	Fl-KKES(**β-Ala**)-CISGDSLISLA	1978.9	22.8 ± 0.9	1.0 ± 0.05	2.3
pep**5**	Fl-KKES(**Sar**)-CISGDSLISLA	1978.9	15.5 ± 0.7	2.1 ± 0.2	3.3
pep**6**	Fl-**SEF**SG-CISGDSLISLA**SR**	2185.9	24.8 ± 0.7	0.7 ± 0.1	1.7
pep**7**	Fl-KKE**A**(**Sar**)-CISGDSLISLA	1962.9	nd[b]	nd[b]	–
pep**8**	Fl-KKESG-(**Hcy**)ISGDSLISLA	1978.8	17.8 ± 0.8	1.2 ± 0.05	2.1
pep**9**	Fl-KKESG-(D-**Cys**)ISGDSLISLA	1964.8	21.2 ± 0.9	0.5 ± 0.2	1.1

[a]The hyphen indicates the peptide bond at the $Ex^N(-1)$-$Int^N(1)$ junction. Fl is 5,6-carboxyfluorescein. Changes relative to pep**1** are in bold. Unnatural amino acids are given in three-letter code and in parentheses. [b]Not determined.

dx.doi.org/10.1021/bi2014823 | *Biochemistry* 2012, 51, 233−242

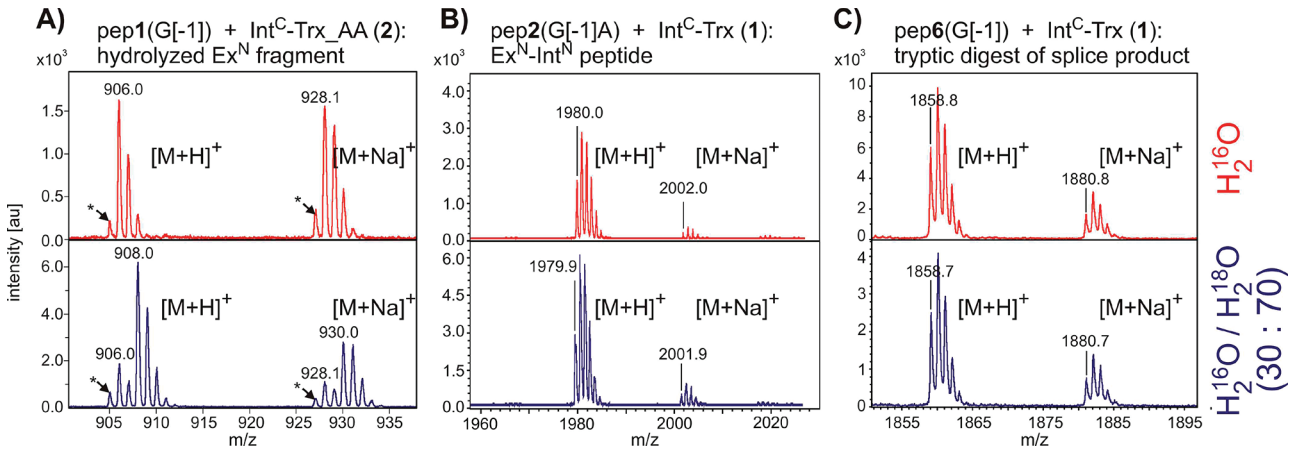


Figure 3. Incorporation of $H_2{}^{18}O$ into reactive intermediates of the N−S acyl shift. The indicated intein constructs were incubated either in $H_2{}^{16}O$ (top) or in an $H_2{}^{16}O/H_2{}^{18}O$ mixture (30:70, bottom), and the fragments containing Ex^N were subsequently analyzed by MALDI-TOF MS. Observed masses are indicated. Only the reaction in panel A showed $H_2{}^{18}O$ incorporation at the 30:70 ratio. Calculated masses are as follows: (A) Fl-KKESG-^{16}OH, $[M + H]^+_{calc}$ = 906.3 Da, $[M + Na]^+_{calc}$ = 928.3 Da; Fl-KKESG-^{18}OH, $[M + H]^+_{calc}$ = 908.3 Da, $[M + Na]^+_{calc}$ = 930.2 Da; (B) pep**2**, $[M + H]^+_{calc}$ = 1979.9 Da, $[M + Na]^+_{calc}$ = 2001.9 Da; (C) Fl-SEFS^{16}GSIEGSGGGSDK, $[M + H]^+_{calc}$ = 1858.6 Da, $[M + Na]^+_{calc}$ = 1880.6 Da. Asterisks denote peptide impurity.

Table 2. Processing of Peptides by Int^C-Trx(1)

peptide	amino acids at the Ex^N(−1)-Int^N(1) junction	splice product formation after 24 h (%)[a]	total conversion of **1** after 24 h (%)[a]
pep**1**	Gly-Cys	100	100
pep**2**	Ala-Cys	7 ± 1	24 ± 1
pep**3**	D-Ala-Cys	26 ± 5	32 ± 2
pep**4**	*β*-Ala-Cys	38 ± 10	92 ± 7
pep**5**	Sar-Cys	37 ± 2	79 ± 16
pep**8**	Gly-Hcy	7 ± 2	17 ± 1
pep**9**	Gly-D-Cys	0	51 ± 16

[a]Relative to reaction with pep**1**; the total conversion of **1** is the sum of its consumption by protein *trans*-splicing and *trans*-cleavage.

We next investigated the ability of the intein to form the thioester at the upstream scissile bond with the various amino acids at position −1. Toward this end, the double mutant Int^C-Trx_AA (construct **2**) was used because it blocks all reactions of the splicing pathway beyond the initial N−S acyl shift. As previously described and mentioned above, with Ala(−1) of pep**2** the oxythiazolidine anion intermediate collapses into the thiazoline ring, resulting in a stable linkage between Ex^N and Int^N in the Ex^N-Int^N/Int^C-Ex^C complex, which can be observed on SDS gels when the sample is not boiled before being loaded.[17] In contrast, the thioester formed with wild-type Gly(−1) of pep**1** quickly hydrolyzes to give a nonfluorescent Int^N/Int^C-Ex^C complex. Figure 5B also shows that the peptides with D-Ala, *β*-Ala, and Sar (pep**3**−pep**5**, respectively) supported the formation of a stable complex at levels similar to that of reference peptide pep**1**. These complexes were all non-fluorescent, thus indicating thioester formation and hydrolysis. Cleavage into the Ex^N and Int^N pieces was also confirmed by MALDI-TOF MS analysis (data not shown, and see below for specific observations with pep**5**). These observations suggest that incomplete thioester formation is not the major cause for the lower yields in the splicing reaction for these peptides.

Together, these findings support the notion that the single methyl group at position −1 specifically in the *S* configuration causes a steric clash, likely with the catalytic His73, and thereby prevents protein splicing, whereas other changes in the backbone structure of the residue at position −1 flanking the active site are quite well tolerated.

Changes at Position −1 Have a Minimal Impact on Intein Fragment Association. The previous results suggested that variations in position −1 mostly affected catalysis and not the formation of the intein complex. For a quantification of Int^N and Int^C fragment binding, we turned to a spectroscopic assay.[17] The binding of a fluorescently labeled peptide to the larger protein results in an increase in fluorescence anisotropy. For this purpose, Int^C-Trx_AAA (construct **3**) with the additional His73Ala mutation was used to block the intein at the stage of fragment association and prevent thioester formation and subsequent hydrolysis. The measured kinetic parameters K_d and k_{on} as well as the calculated k_{off} values are summarized in Table 1. The values of the peptides containing D-Ala, *β*-Ala, and Sar differed only slightly

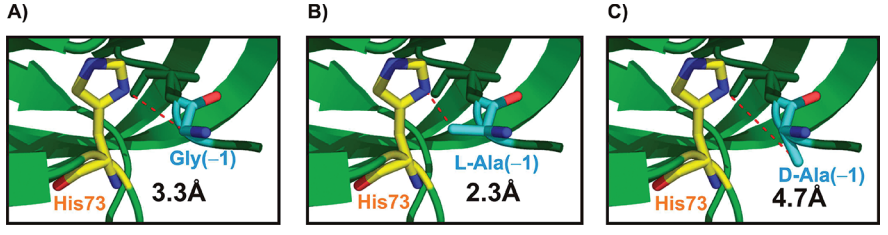


Figure 4. Distance between His73 and selected atoms of the amino acid at position −1 in the *Ssp* DnaB intein. These images were created using PyMol and Protein Data Bank entry 1MI8. (A) Crystal structure of the intein (including the C1A and N154A mutations). (B and C) L-Ala and D-Ala, respectively, were modeled into the active site using PyMol.

dx.doi.org/10.1021/bi2014823 | *Biochemistry* 2012, 51, 233−242

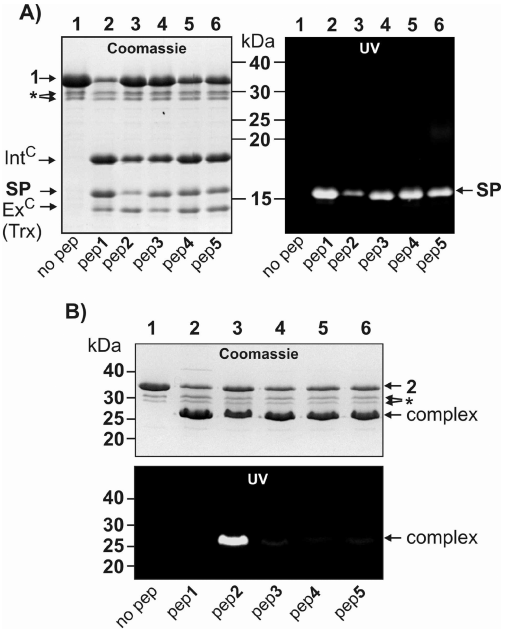


Figure 5. Influence of amino acid substitutions at position −1. Shown are SDS−PAGE gels under UV illumination or stained with Coomassie brilliant blue. The reactions were performed for 24 h. (A) Protein *trans*-splicing. (B) Intein complex formation investigated without boiling the protein samples in SDS buffer. The complex exhibits a different migration behavior compared to that of free construct **2**. SP denotes the splice product; the asterisk denotes protein impurities.

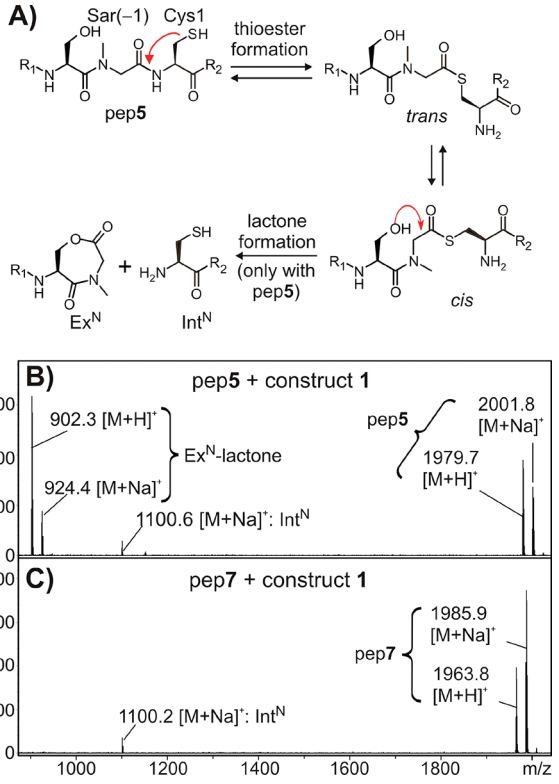


Figure 6. Aberrant cleavage of the thioester induced by Sar(−1). (A) Proposed model for the attack of the Ser(−2) hydroxyl side chain on the thioester favored by the higher *cis* content of the Sar(−1)-Cys1 peptide bond. (B and C) MALDI-TOF MS analyses of the reactions with the indicated intein fragments. The Ex^N-lactone was not observed when Ser(−2) was replaced with Ala. Observed masses are indicated. Calculated masses are as follows: Ex^N-lactone, $[M + H]^+_{calc}$ = 902.4 Da, $[M + Na]^+_{calc}$ = 925.3 Da; Int^N, $[M + H]^+_{calc}$ = 1078.5 Da, $[M + Na]^+_{calc}$ = 1100.5 Da; pep**5**, $[M + H]^+_{calc}$ = 1979.9 Da, $[M + Na]^+_{calc}$ = 2001.9 Da; pep7, $[M + H]^+_{calc}$ = 1963.9 Da, $[M + Na]^+_{calc}$ = 1985.9 Da.

from those containing Gly or Ala at position −1. Thus, changes at this position indeed have only little impact on intein fragment binding. Therefore, the observed functional differences must mostly stem from the inability of the intein to subsequently adopt the precise alignment for catalysis or from the kinetics of the individual steps of the splicing reaction.

Identification of an Alternative Intramolecular Cyclization Induced by Sarcosine at Position −1. We observed a peculiarity when analyzing the reaction of the peptide with sarcosine at position −1 (pep**5**) with construct **2**. By MS analysis, we detected a fragment that corresponded to the Ex^N peptide as a result of a cleavage at the scissile bond; however, for a hydrolysis product, its mass was too small by 18 Da (data not shown). Interestingly, the same fragment ($[M + H]^+$ = 902.3 Da) could be detected in splicing reactions of pep**5** with construct **1** (see Figure 6B). We hypothesized that a loss of water through a cyclization event might account for this difference. Sarcosine, like other N-alkylated amino acids such as proline, less strongly disfavors the *cis* over *trans* conformation in the aa-Sar peptide bond compared to regular peptide bonds. We therefore wondered whether the more populated *cis* conformation of the Ser-Sar peptide bond in the Ex^N sequence could position the *β*-hydroxyl group of Ser(−2) into a privileged position for a nucleophilic attack on the Ex^N thioester to give the seven-membered lactone (see Figure 6A).

To probe this theory, we synthesized pep**7** with Ser(−2) replaced with Ala (see Table 1). In the *trans*-splicing assay with construct **1**, the semisynthetic splice product was formed with an efficiency comparable to that with pep**5** (data not shown); however, the mass of the cyclized Ex^N could no longer be detected (see Figure 6C). Likewise, this product was not observed upon incubation of pep**7** with construct **2** in the thioester formation assay. These observations strongly support the proposed reaction scheme for the formation of the seven-membered lactone side product (Figure 6A).

Unnatural Mutation Analysis of the Catalytic Cys1 Residue. We next focused on the catalytic Cys1 residue involved in thioester formation at the upstream splice junction. Substitution of Cys1 with unnatural amino acids could help in studying the importance of the thiol side chain positioning for protein splicing in general and for the initial N−S acyl shift in particular. Two peptides containing homo-Cys (Hcy, pep**8**) and D-Cys (pep**9**) at position 1 were synthesized (see Table 1). Both peptides were subjected to splicing reactions with construct **1**, which were then analyzed by SDS−PAGE as discussed above. The reaction with pep**9** showed no splicing, only the formation of the C-terminal cleavage product Trx-His_6. However, the reaction with pep**8** still yielded a small amount of splice product, ∼7% relative to the amount for the reaction with pep**1** (Figure 7A and Table 2). Determination of the K_d and k_{on} values for the intein fragment association using fluorescence anisotropy spectroscopy revealed only slight changes compared to those of reference peptide pep**1** with a Cys1 residue (Table 1), indicating that the decrease or loss of splicing activity was not caused by altered binding affinities. To identify the step in the splicing pathway at which the Hcy and D-Cys residues were detrimental to high product yields, we analyzed if peptides pep**8** and pep**9** still supported thioester

dx.doi.org/10.1021/bi2014823 | *Biochemistry* 2012, 51, 233−242

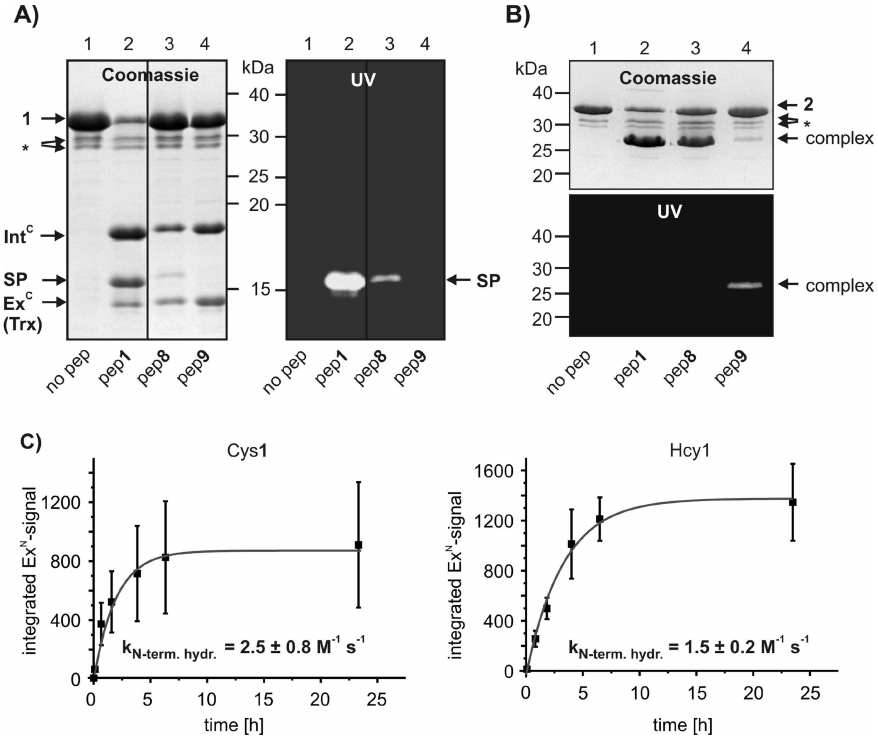


Figure 7. Substitution of the catalytic Cys1 with Hcy and D-Cys. Shown are SDS−PAGE gels for analyzing the protein *trans*-splicing reaction using Int^C-Trx (construct **1**) (A) as well as the complex and thioester formation using Int^C-Trx_AA (construct **2**) (B). Samples in panel B were not boiled before being loaded onto a gel to preserve the noncovalent intein complex. The complex exhibits a migration behavior different from that of the free construct **2**. The reactions were performed for 24 h. In panels A and B, SP denotes the splice product and asterisks denote protein impurities. (C) Time course of N-terminal cleavage through thioester hydrolysis in reactions with pep**8** or pep**9** and Int^C-Trx_AA (construct **2**).

formation using the gel shift assay described above with protein Int^C-Trx_AA (construct **2**). Interestingly, pep**9** with D-Cys led to only small amounts of the intein complex, but this complex was highly fluorescent, indicating that the fluoresceine-tagged Ex^N sequence was not cleaved off by thioester hydrolysis, likely because the thioester could not be formed (Figure 7B). In contrast, pep**8** with Hcy supported formation of the intein complex at levels approaching the reaction with reference peptide pep**1**, but this complex was completely nonfluorescent under UV light. This observation showed that thioester formation must have taken place, followed by its hydrolysis. Thus, whereas D-Cys completely blocked thioester formation, Hcy supported this N−S acyl shift. To measure the rates of thioester formation with the Cys- and Hcy-containing peptides, we assumed that the hydrolysis was fast compared to thioester formation and established an HPLC assay to monitor the thioester hydrolysis rate (data not shown). The plotted data were fit to second-order rate constants for thioester hydrolysis to give 2.5 ± 0.8 M^{-1} s^{-1} for pep**1** (Cys) and 1.5 ± 0.2 M^{-1} s^{-1} for pep**8** (Hcy) (Figure 7C). Thus, formation and hydrolysis of the Hcy thioester are not significantly slower than those of the Cys thioester. Taken together, these results suggest that in the case of Hcy the splicing pathway is impaired after the N−S acyl shift.

Finally, we tested the ability of the previously reported evolved Int^{CM86} mutant[20] in the same protein context (construct **1a**) to splice with Hcy at the catalytic position. Figure 8 shows that significantly higher splice product yields compared to that of construct **1** were obtained. The evolved mutant **1a** was converted to 52% after 24 h (compared to 13% for **1**), of which 43% led to splice product formation (compared to 17% for **1**).

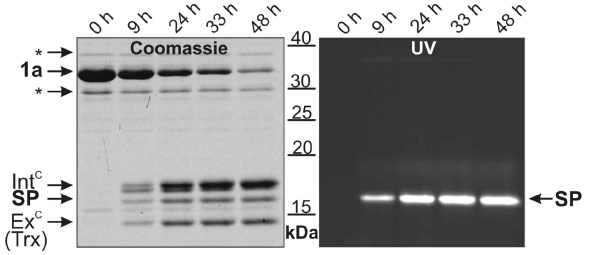


Figure 8. Evolved intein mutant capable of splicing with an unnatural amino acid as the catalytic residue. The evolved M86 mutant of the *Ssp* DnaB intein (Int^{CM86}-Trx, construct **1a**) was incubated with pep**8** containing Hcy at catalytic position 1 of the N-terminal splice junction. Shown are SDS−PAGE gels. Note that the Int^C fragment gives rise to two bands that correspond to the intact C-terminal succinimide and its hydrolyzed form. SP denotes the splice product; asterisks denote protein impurities.

DISCUSSION

Inteins catalyze as single-turnover enzymes a remarkable sequence of acyl rearrangements that ultimately effect their excision out of a precursor protein and the ligation of the flanking extein sequences with a native peptide bond. While the general pathway of protein splicing is well understood, many open questions remain about how the individual steps are catalyzed by the intein and how these are coordinated in space and time to avoid unproductive side reactions. Likewise, it has been documented for many inteins that certain immediately flanking amino acids inhibit or slow the protein splicing reaction, but the reasons for this at the atomic level are unknown. Understanding this structure−function relationship will be crucial for the full exploitation of the tremendous

dx.doi.org/10.1021/bi2014823 | *Biochemistry* 2012, 51, 233−242

potential of inteins in many applications, e.g., in biotechnology, biomedicine, and basic protein research.

In this study, we have taken advantage of a previously developed semisynthetic split *Ssp* DnaB intein.[16,24] In this intein, the N-terminal scissile bond including the flanking N-extein residues as well as the Cys1 residue of the catalytic site are located on the short Ex^N-Int^N peptide. All residues that are directly involved in the initial N−S acyl shift for thioester formation are thus accessible to precise manipulation by organic peptide synthesis. Whereas studies using recombinant proteins can only introduce genetically encoded amino acids to interrogate the structure−function relationship in the region, our approach allows for the introduction of unnatural building blocks of virtually arbitrary structure.

A previous observation was one of the motivations for this study. The semisynthetic split *Ssp* DnaB intein was unable to splice efficiently with the Gly(−1)Ala substitution and promoted thiazoline ring formation at the N-terminal splice junction in the context of the double mutant Int^C-Trx_AA (construct **2**).[17] This was a new byproduct formed from the oxythiazolidine anion intermediate in the protein splicing pathway. The occurrence of such an aberrant reaction as a consequence of slight deviations of the wild-type sequence, as observed earlier also for the N- and C-terminal cleavage reactions, highlighted the importance of the precisely aligned residues in the inteins' active site for efficient protein splicing. While mutations, of course, may have disadvantageous consequences, it was surprising to us that a single extra methyl group could have such a dramatic impact. In this work, we sought to understand the effect of the Gly(−1)Ala mutation in more detail. We first performed back-incorporation assays with $H_2{}^{18}O$ that excluded the formal possibility that the elimination of water to form the thiazoline might be a reversible and regular reaction, also in the wild-type intein. We then introduced other subtle structural changes by adding the "additional methyl group" at other immediately adjacent positions. In fact, all these substitutions, i.e., Gly(−1) changed to Sar, *β*-Ala, and D-Ala, supported significantly higher protein splicing yields than the Gly(−1)Ala mutation and did not lead to the formation of the thiazoline byproduct in combination with the double mutant Int^C-Trx_AA. These findings show that on the one hand a single methyl group can have a dramatic impact, probably because of a steric clash of Ala(−1) with His73 that prevents correct organization of the active site. On the other hand, such a single methyl group may be well tolerated. This is noteworthy because the change of Gly(−1) to Sar or *β*-Ala should also affect hydrogen bonding networks, as in these cases the amide proton is eliminated or the entire upstream peptide bond of this residue is repositioned by the distance of one extra methylene group, respectively. Interestingly, the constants for binding of each of these Ex^N-Int^N peptides to the Int^C-Ex^C construct were also very similar to that of the reference peptide with Gly(−1). Together, these results suggest that the residues located directly upstream of the intein do not have a significant impact on intein fragment association, intein folding, or intein activity, unless a specific constraint is created that negatively affects folding of the active site.

The finding that the intein spliced significantly better with D-Ala at position −1 compared to the genetically encoded stereoisomer also led us to consider the possibility that a twisting of the native Gly(−1) into the conformational space of D-amino acids is required for thioester formation. Indeed, a further inspection of the crystal structure of the *Ssp* DnaB intein revealed that Gly(−1) adopts unusual dihedral angles of −153.6° (Ψ) and −82° (Φ) that lie in a Ramachandran plot outside the region typically favored by L-amino acids. This may indicate that a strained conformation has to be adopted; however, these dihedral angles do not correspond to the region that would be preferred by D-amino acids. The fact that the crystal structure represents a mutant inactive in protein splicing further complicates the interpretation of these data. In this context, it is also of interest to note that we have recently isolated an evolved mutant of the *Ssp* DnaB intein, mutant M86, that exhibited more general splicing capabilities with respect to the flanking amino acids.[20] In particular, the derived semisynthetic split intein variant was able to efficiently splice with the Gly(−1)Ala substitution, which further corroborates the model in which a particular extein structure is not a prerequisite for splicing, but rather that a robust intein can impose its folding requirements on an unfavorable extein structure.

The thioester at the upstream splice junction is a very reactive species. Inteins must have evolved to exclude water molecules from their active sites that might hydrolyze off the N-extein fragment from this intermediate. Our studies with sarcosine at position −1 have revealed yet another potential side reaction of the thioester. Presumably because of the higher *cis* content of the peptide bond between Ser(−2) and Sar(−1), the Ser hydroxyl side chain was able to attack the thioester to give rise to the seven-membered lactone (see Figure 6). This reaction was not observed when Ser(−2) was substituted with Ala. Interestingly, native inteins that operate with proline as the last amino acid of the N-extein are known. One would expect that these inteins have undergone evolutionary pressure to prevent a similar reaction caused by the higher *cis* content of the Xaa-Pro peptide bond. A survey of the known inteins listed in InBase[9] indeed confirms this hypothesis, as none of the 40 known inteins with Pro(−1) has a Ser, Thr, or Cys at position −2. However, more intein sequences would be necessary for a statistically more significant statement, as 38 of these 40 inteins belong to the homologous ribonucleotide reductase inteins of classes I−III.

To address the importance of the active site thiol group orientation, we replaced the catalytic Cys1 with the non-proteinogenic D-Cys and Hcy. The finding that the peptide with D-Cys did not support any splicing or thioester formation is in line with the notion that the active site is not configured for the inverted stereochemistry at this critical amino acid. Interestingly, the additional methylene group in the Hcy side chain revealed another twist of a subtle structural change. In this case, the change had only little impact on the first reaction in which this residue is involved, i.e., the N−S acyl shift in thioester formation; however, protein *trans*-splicing was severely inhibited. Efficient thioester formation, presumably through the six-membered intermediate, raises a question about the role of the intein in catalyzing this step. Hcy is known to be competent in the chemical equivalent of the reverse reaction, i.e., native chemical ligation,[25] with only 2-fold slower reaction kinetics,[26] and thus very similar to Cys in these acyl shift reactions. However, a similar reaction rate in the active site of the intein is surprising, if one assumes that deprotonation of the Cys1 side chain thiol and stabilization of the thiolate anion should be assisted by acid−base catalysis. A base optimally positioned for abstraction of a proton from Cys1 would be expected to be misaligned for Hcy1. Furthermore, intein-mediated deprotonation of Hcy would appear to be even more

dx.doi.org/10.1021/bi2014823 | *Biochemistry* 2012, 51, 233−242

important for efficient catalysis when one considers the pK_a of its thiol group, which is reported to be 0.5−2 units higher than that of Cys, depending on the reference.[27,28] Therefore, our data suggests that a base-mediated deprotonation of the Cys1 thiol group is not part of the catalytic strategy of inteins, at least not of this particular intein, or at least it is clearly not the rate-limiting step.

This piece of evidence adds to a long and still unsolved discussion of how the N−S acyl shift is mediated by the intein. Indeed, a general base that deprotonates Cys1 has not yet been clearly identified. In crystal structures, no amino acid side chain capable of abstracting a proton is properly positioned to act as a base. Consistent with these observations is a study by Paulus and co-workers that reported a pK_a of 8.2 for Cys1 in the *Mtu* RecA intein, indicating that no base is involved in its activation.[29] On the other hand, a recent study by Wang and Belfort concluded from pK_a titrations by NMR and QM/MM calculations that the highly conserved block B histidine in the *Mtu* RecA intein (corresponding to His73 in the *Ssp* DnaB intein) first acts as a base to deprotonate the thiol group of Cys1 and then as an acid to protonate the liberated α-amino group.[13] Also on the basis of pK_a titrations by NMR and a mutagenesis study, Wang and co-workers suggest that Asp422 of block F in the *Mtu* RecA intein has an elevated pK_a and thereby stabilizes in its protonated form the cysteine thiolate before the attack on the peptide bond.[30] However, also for the corresponding D136 of the *Ssp* DnaB mini-intein,[22,31] one would not expect an optimal positioning to stabilize the thiolate of the Hcy side chain in the context of our semisynthetic split intein. Furthermore, these studies do not explain how a hydroxyl group would be deprotonated in the case of those inteins that employ a Ser1 residue, as a much stronger base should be necessary.

An alternative strategy for rearranging the scissile peptide bond into the thioester seems conceivable when insights from recent developments for the synthesis of peptide thioesters are taken into account.[32,33] Peptide bonds at Cys residues undergo acyl shift reactions in an equilibrium from which the thioester can be obtained by using a cysteinyl prolyl ester unit to lock the liberated amino group in a diketopiperazine structure.[34,35] Thus, a major role of the intein could be to stabilize the thioester once is has been formed in an equilibrium reaction. To promote the forward reaction, an active thiol deprotonation might not be required. Earlier studies suggested a ground-state destabilization of the peptide bond as an alternative or additional strategy.[13,15,36] Twisting the scissile peptide bond into a conformation that favors formation of the oxythiazolidine anion intermediate would force the amide nitrogen closer to sp^3 hybridization such that its elevated pK_a could more easily accept a proton.[37]

Why is the Hcy thioester not efficiently processed to the splice product? There are two possible reasons: because the extra methylene group the Hcy thioester is not properly positioned for the attack of the downstream Ser(+1), or the Hcy thioester cannot support a conformational change required for the trans-esterification step. These explanations would underline once more the importance of the perfectly aligned geometry in the active site of the intein. Strikingly, we find that the M86 mutant of the *Ssp* DnaB intein evolved for higher splicing efficiency and extein promiscuity[20] also exhibits a significantly better ability to process the Hcy thioester to the splice product (Figure 8). This finding is surprising because the mutant intein was not evolved for splicing with this unnatural amino acid substitution in its active site. Possible explanations for this behavior are, for example, that the generally better kinetic performance of the M86 mutant is able to partly compensate for the decreased reaction velocity caused by the misaligned thioester, i.e., a faster transesterification reaction on a bad substrate would lead to higher product yields, and that a better packed active site also improves processing of the unnatural intermediate.

In summary, our results show that subtle structural alterations at the upstream extein-intein junction can either be well tolerated or significantly impair the protein splicing pathway. Characterizing and understanding these dependencies will be important for the future exploitation of inteins for various applications.

AUTHOR INFORMATION

Corresponding Author

*E-mail: henning.mootz@uni-muenster.de. Telephone: +49-251-83-33005. Fax: +49-251-83-33007.

Present Addresses

[†]Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Str. 4, 72076 Tübingen, Germany.

[‡]Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Wolfgang-Pauli-Str. 16, 8093 Zürich, Switzerland.

Author Contributions

D.S. and C.L. contributed equally to this work.

Funding

Funding was kindly provided by the Deutsche Forschungsgemeinschaft (DFG Grants MO1073/3-1 and SCHW1163/2-1) and the Fonds der Chemischen Industrie. C.L. was a recipient of a Ph.D. stipend from the Fonds der Chemischen Industrie. I.V.T. acknowledges her Ph.D. stipend from the German National Merit Foundation.

ACKNOWLEDGMENTS

We thank Gerrit Volkmann for discussions and comments on the manuscript.

ABBREVIATIONS

AA, N154A and S(+1)A mutations in construct **2**; AAA, H73A, N154A, S(+1)A mutations in construct **3**; Ex^N and Ex^C, N- and C-terminal exteins, respectively; Hcy, homocysteine; His_6, hexahistidine tag; Int^N and Int^C, N- and C-terminal intein halves, respectively, of the split *Ssp* DnaB intein; Int^{CM86}, evolved mutant of the C-terminal intein half of the split *Ssp* DnaB intein; Sar, sarcosine; *Ssp* DnaB, intein of the DnaB helicase of *Synechocystis* spp. PCC6803; Trx, thioredoxin.

REFERENCES

(1) Noren, C. J., Wang, J., and Perler, F. B. (2000) Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew. Chem., Int. Ed. 39*, 450−466.

(2) Mills, K. V., and Perler, F. B. (2005) The mechanism of inter-mediated protein splicing: Variations on a theme. *Protein Pept. Lett. 12*, 751−755.

(3) Paulus, H. (2000) Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem. 69*, 447−496.

(4) Evans, T. J. T., and Xu, M. Q. (2002) Mechanistic and kinetic considerations of protein splicing. *Chem. Rev. 102*, 4869−4884.

(5) Mootz, H. D. (2009) Split inteins as versatile tools for protein semisynthesis. *ChemBioChem 10*, 2579−2589.

(6) Vila-Perello, M., and Muir, T. W. (2010) Biological applications of protein splicing. *Cell 143*, 191−200.

(7) Volkmann, G., and Iwai, H. (2010) Protein *trans*-splicing and its use in structural biology: Opportunities and limitations. *Mol. BioSyst. 6*, 2110−2121.
(8) Perler, F. B. (2005) Protein splicing mechanisms and applications. *IUBMB Life 57*, 469−476.
(9) Perler, F. B. (2002) InBase: The Intein Database. *Nucleic Acids Res. 30*, 383−384.
(10) Anraku, Y., Mizutani, R., and Satow, Y. (2005) Protein splicing: Its discovery and structural insight into novel chemical mechanisms. *IUBMB Life 57*, 563−574.
(11) 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.
(12) Southworth, M. W., Benner, J., and Perler, F. B. (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J. 19*, 5019−5026.
(13) Du, Z., Shemella, P. T., Liu, Y., McCallum, S. A., Pereira, B., Nayak, S. K., Belfort, G., Belfort, M., and Wang, C. (2009) Highly conserved histidine plays a dual catalytic role in protein splicing: A pKa shift mechanism. *J. Am. Chem. Soc. 131*, 11581−11589.
(14) Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R. Jr., and Sacchettini, J. C. (1998) Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat. Struct. Biol. 5*, 31−36.
(15) Romanelli, A., Shekhtman, A., Cowburn, D., and Muir, T. W. (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proc. Natl. Acad. Sci. U.S.A. 101*, 6397−6402.
(16) 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. 45*, 5218−5221.
(17) Ludwig, C., Schwarzer, D., and Mootz, H. D. (2008) Interaction studies and alanine scanning analysis of a semi-synthetic split intein reveal thiazoline ring formation from an intermediate of the protein splicing reaction. *J. Biol. Chem. 283*, 25264−25272.
(18) Binschik, J., Zettler, J., and Mootz, H. D. (2011) Photocontrol of protein activity mediated by the cleavage reaction of a split intein. *Angew. Chem., Int. Ed. 50*, 3249−3252.
(19) Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y., and Satow, Y. (2002) Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides. *J. Mol. Biol. 316*, 919−929.
(20) 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. 39*, 34440−34447.
(21) Ludwig, C., Schwarzer, D., Zettler, J., Garbe, D., Janning, P., Czeslik, C., and Mootz, H. D. (2009) Semisynthesis of proteins using split inteins. *Methods Enzymol. 462*, 77−96.
(22) Ding, Y., Xu, M. Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J. Biol. Chem. 278*, 39133−39142.
(23) Li, A. J., and Nussinov, R. (1998) A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins 32*, 111−127.
(24) Sun, W., Yang, J., and Liu, X. Q. (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J. Biol. Chem. 279*, 35281−35286.
(25) Tam, J. P., and Yu, Q. (1998) Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones. *Biopolymers 46*, 319−327.
(26) Kennan, A. J., Haridas, V., Severin, K., Lee, D. H., and Ghadiri, M. R. (2001) A de novo designed peptide ligase: A mechanistic investigation. *J. Am. Chem. Soc. 123*, 1797−1803.
(27) Benesch, R. E., and Benesch, R. (1955) The acid strength of the -SH group in cysteine and related compounds. *J. Am. Chem. Soc. 77*, 5877.
(28) Friedman, M., Cavins, J. F., and Wall, J. S. (1965) Relative nucleophilic reactivities of amino groups and mecaptide ions in addition reactions with α,β-unsaturated compounds. *J. Am. Chem. Soc. 87*, 3672−3682.
(29) Shingledecker, K., Jiang, S., and Paulus, H. (2000) Reactivity of the cysteine residues in the protein splicing active center of the *Mycobacterium tuberculosis* RecA intein. *Arch. Biochem. Biophys. 375*, 138−144.
(30) Du, Z., Zheng, Y., Patterson, M., Liu, Y., and Wang, C. (2011) pK_a coupling at the intein active site: Implications for the coordination mechanism of protein splicing with a conserved aspartate. *J. Am. Chem. Soc. 133*, 10275−10282.
(31) Wu, H., Xu, M. Q., and Liu, X. Q. (1998) Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim. Biophys. Acta 1387*, 422−432.
(32) Kang, J., and Macmillan, D. (2010) Peptide and protein thioester synthesis via N → S acyl transfer. *Org. Biomol. Chem. 8*, 1993−2002.
(33) Mende, F., and Seitz, O. (2011) 9-Fluorenylmethoxycarbonyl-based solid-phase synthesis of peptide α-thioesters. *Angew. Chem., Int. Ed. 50*, 1232−1240.
(34) Kawakami, T., and Aimoto, S. (2007) Sequential peptide ligation by using a controlled cysteinyl prolyl ester (CPE) autoactivating unit. *Tetrahedron Lett. 48*, 1903−1905.
(35) Zanotti, G., Pinnen, F., and Lucente, G. (1985) Cyclization under mild conditions of cysteine containing peptides. *Tetrahedron Lett. 26*, 5481−5484.
(36) Mills, K. V., and Paulus, H. (2005) Biochemical mechanisms of intein-mediated protein splicing. In *Homing Endonucleases and Inteins* (Belfort, M., Berbyshire, V., Stoddard, B. L., and Wood, D. W., Eds.) pp 233−255, Springer-Verlag, Berlin.
(37) Johansson, D. G., Wallin, G., Sandberg, A., Macao, B., Aqvist, J., and Hard, T. (2009) Protein autoproteolysis: Conformational strain linked to the rate of peptide cleavage by the pH dependence of the N → O acyl shift reaction. *J. Am. Chem. Soc. 131*, 9475−9477.