Published in final edited form as:

Biotechnol Bioeng. 2019 April ; 116(4): 709–721. doi:10.1002/bit.26875.

Structure of an Engineered Intein Reveals Thiazoline Ring and Provides Mechanistic Insight

C. Seth Pearson¹, Reza Nemati², Binbin Liu³, Jing Zhang³, Matteo Scalabrin², Zhong Li³, Hongmin Li^{3,4}, Dan Fabris², Marlene Belfort^{4,5,*}, and Georges Belfort^{1,*}

¹Howard P Isermann Department of Chemical and Biological Engineering and Center for Biotechnology and Interdisciplinary Studies, RPI, Troy, NY 12180

²Department of Chemistry, State University of New York at Albany, Albany, NY 12222

³Division of Computational & Structural Biology, Wadsworth Center, NY State Dept. of Health, Albany, NY 12208

⁴Department of Biomedical Sciences, School of Public Health, State University of New York at Albany, Albany NY, 12201

⁵Department of Biological Sciences, State University of New York at Albany, Albany, NY 12222

Abstract

We have engineered an intein which spontaneously and reversibly forms a thiazoline ring at the native N-terminal Lys-Cys splice junction. We identified conditions to stabilize the thiazoline ring and provide the first crystallographic evidence, at 1.54 Å resolution, for its existence at an intein active site. The finding bolsters evidence for a tetrahedral oxythiazolidine splicing intermediate. Additionally, the pivotal mutation maps to a highly conserved B-block threonine, which is now seen to play a causative role not only in ground-state destabilization of the scissile N-terminal peptide bond, but also in steering the tetrahedral intermediate toward thioester formation, giving new insight into the splicing mechanism. We demonstrate the stability of the thiazoline ring at neutral pH as well as sensitivity to hydrolytic ring opening under acidic conditions. A pH cycling strategy to control N-terminal cleavage is proposed, which may be of interest for biotechnological applications requiring a splicing activity switch, such as for protein recovery in bioprocessing.

Keywords

Protein splicing; cleavage control; thiazoline crystal structure; molecular switch

INTRODUCTION

Inteins are protein elements that autocatalytically excise themselves from the host protein in a process called protein splicing. The two flanking sequences, called the N- and C-exteins, are subsequently ligated with a native peptide bond, resulting in a functional mature protein (Hirata et al. 1990; Kane et al. 1990). Insights into the mechanism of protein splicing,

*Co-corresponding authors: mbelfort@albany.edu and belfog@rpi.edu.

specifically the roles of conserved domains and residues, have allowed engineering of inteins into biotechnology tools. These tools utilize strategies such as decoupling the splicing reaction to produce self-cleaving inteins for purification processes (Banki et al. 2005; Chong et al. 1997; Ge et al. 2005; Muir et al. 1998; Perler and Adam 2000; Southworth et al. 1999; Wood et al. 2000; Wood et al. 1999), replacing an endonuclease domain, not required for splicing, with a receptor domain for small molecule splicing activation (Buskirk et al. 2004; Skretas and Wood 2005), generating ester intermediates for expressed protein ligation (Evans et al. 1998; Muir et al. 1998), and disulfide-trapping a catalytic cysteine to generate a redox sensor (Callahan et al. 2013), among numerous other applications (Shah and Muir 2014; Topilina and Mills 2014). Although impressive strides have been made, further insights into catalytic control are required to expand the applications for inteins as biotechnology tools.

Inteins typically have several conserved motifs or blocks (A, B, F, and G) which contain the residues that are largely responsible for mediating the autocatalytic splicing reaction (Perler et al. 1997; Pietrokovski 1994; Pietrokovski 1998) (Fig 1A). These sequences often flank an endonuclease domain that is not involved in splicing. Block A occurs at the N-terminus of the intein and contains a cysteine, serine or threonine residue in the 1 position which performs an essential nucleophilic displacement (Pereira et al. 2011; Perler and Adam 2000; Saleh and Perler 2006). The B block is somewhat distant in primary sequence, but is close to the A block spatially, and contains a highly conserved TxxH motif. The F block, again close spatially to the active site, contains a highly conserved aspartate, whereas the G block consists of the C-terminus of the intein which contains a highly conserved penultimate histidine and terminal asparagine. There is an adjacent cysteine, serine or threonine residue in the extein at the +1 position, which is also involved in an essential nucleophilic displacement.

The first step of intein splicing consists of an N-X acyl shift involving cysteine, serine or threonine. In the RecA intein, which splices by the canonical pathway, this shift is characterized by a nucleophilic attack by the side chain of the first cysteine of the intein on the carbonyl carbon of the adjacent N-extein residue (Xu and Perler 1996). This attack is postulated to result in a tetrahedral intermediate. The nitrogen of this intermediate is then protonated, which is followed by cleavage of the C-N bond and subsequent formation of a linear thioester, completing the first step (Fig 1B,C) (Mizutani et al. 2002).

An off-pathway reaction was postulated (Ludwig et al. 2008), in which the alkoxide of the tetrahedral intermediate is aberrantly protonated, resulting in cleavage of the C-O bond of the *ssp* DnaB intein and collapse to a dehydrated thiazoline ring (Fig. 1C). However, the authors were unable to obtain unequivocal physical evidence for the thiazoline ring. A follow-up study using the same *ssp* DnaB construct utilized the lack of potential back-incorporation of H₂¹⁸O to demonstrate that thiazoline ring formation is not a reversible step of the standard protein splicing (Schwarzer et al. 2012).

Although aberrant chemistry for an intein, several families of cyclodehydratases exist which catalyze thiazoline ring formation during the production of bioactive small molecules, including antibiotic peptides such as bacitracin and siderophores such as yersiniabactin

(Gehring et al. 1998a; Gehring et al. 1998b; Ishihara and Shimura 1988). The catalytic similarities (formation of an oxythiazolidine intermediate) and differences (resolution into thioester vs thiazoline ring) between these enzymes have been noted (Ludwig et al. 2008; Walsh et al. 2012), but, while much is known about inteins, the atomic details of these cyclodehydratases remain obscure (Kelly 2012).

For standard intein chemistry, numerous mechanistic pathways have been proposed to explain the catalytic activity during the first step of splicing, namely (i) ground state destabilization, (ii) activation of the nucleophilic Cys1 side chain, and (iii) tetrahedral intermediate stabilization. Evidence for each mechanism has been presented based on several different conserved intein residues (Paulus 2000; Saleh and Perler 2006; Xu and Perler 1996).

The B-block histidine (part of the conserved TxxH motif) has been invoked repeatedly as a catalytic factor in the initial N-S acyl shift and has indeed been implicated in all three catalytic mechanisms. This residue is highly conserved and has been shown to be essential for this reaction (Kawasaki et al. 1997; Pietrokovski 1994). It is located within hydrogen bonding distance of the N-terminal splice junction and was initially postulated to protonate the amide nitrogen of the tetrahedral intermediate (Klabunde et al. 1998; Mizutani et al. 2004; Mizutani et al. 2002; Poland et al. 2000).

The B-block threonine of the TxxH motif has been shown to be involved in catalyzing the initial N-S acyl shift. A crystalized trapped precursor of the *Ssp* DnaE intein showed abnormal bond angles between Cys1 and the adjacent peptide bond of the N-extein (Fig. S1), a finding corroborated in crystal structures of the *Mxe* GyrA and *Sce* VMAI inteins (Dearden et al. 2013; Klabunde et al. 1998; Poland et al. 2000). This strain was shown to be relieved by mutation of the threonine. These experimental findings were supported by QM/MM calculations and implicate the threonine in the ground state destabilization mechanism (Dearden et al. 2013). The mechanisms discussed above are not necessarily contradictory and each plays a role in driving the first step of splicing, which will likely vary on an intein-specific basis (Volkman and Liu 2011).

Here, using mass spectrometry, crystallography and mutational studies with a minimized *RecA* intein (Fig. 1A), we provide unequivocal evidence of thiazoline ring formation in an engineered intein, firmly supporting the hypothesis of a tetrahedral intermediate. We also elucidated an additional role of the B-block threonine in directing the fate of the oxythiazolidine intermediate. We further show that the thiazoline ring, which is stable at basic pH, can be opened under acidic conditions, suggesting the utility of this mutation as a switch for modulating protein cleavage.

RESULTS AND DISCUSSION

B Block Threonine Mutant Contains a Thiazoline Ring at N-terminal Splice Junction

In an effort to engineer a cleavage-compromised *Mtu* *RecA* intein that binds divalent metal ions, we introduced two mutations to the previously-described I-SM mini-intein construct from which the endonuclease domain has been deleted (Hiraga et al. 2009): a B-block

threonine (Thr) to alanine (Ala) mutation (T70A) and a terminal asparagine (Asn) to alanine (Ala) mutation (N139A). This construct is termed I-TM. The Asn mutation was proposed to arrest C-terminal cleavage by preventing the essential Asn cyclization step, while the B block Thr was proposed to arrest N-terminal cleavage by eliminating destabilization of the intein junction peptide bond which primes the preceding N-extein carbonyl carbon for nucleophilic attack by Cys1 of the intein.

Expression of the I-TM intein mutant indicated that splicing, N-terminal cleavage, and C-terminal cleavage were arrested (Fig. S2), yet no evidence of metal binding could be detected despite the presence of the putative metal binding residues. Platinum binding has been shown to be mediated mainly by the sidechains of Cys1 and Cys+1 of the RecA intein (Chan et al. 2016; Zhang et al. 2011), suggesting, among other possibilities, that one or both cysteines (Cys) may be perturbed spatially or otherwise modified. Interestingly, the I-TM (T70A/N139A) mutant exhibited the most intense signal at exactly 18 Da less than the expected mass when measured by ESI-MS (monoisotopic $M_{\text{obs}} = 16,454.60$ Da, Fig. 2). Thus, the expected monoisotopic mass calculated from the sequence without the thiazoline ring (M_{calc} of 16,472.73 Da) was not detected.

It should be noted that the intein sprays as a major peak with several smaller peaks at -28 Da intervals, which arise from the occasional misincorporation of a lysine (Lys) residue for an arginine (Arg) residue during protein overexpression. The *Mtu* RecA intein contains several Arg repeats which utilize the rare codon AGA. It has previously been shown that *Escherichia coli* is prone to misincorporate Lys for Arg where this rare codon is used, especially in repeat regions (Calderone et al. 1996; You et al. 1999). However, as these misincorporations occur in distal regions from the active site they have no observable functional importance (You et al. 1999).

Because of the -18 Da discrepancy and the likely participation of a Cys residue, we suspected a dehydration reaction involving Cys1, as previously observed in the *Ssp* DnaB intein (Ludwig et al. 2008). We therefore determined a crystal structure of the I-TM T70A/N139A mutant. The 1.54 Å resolution structure revealed a thiazoline ring at the N-extein-intein junction between the native Lys-1 and Cys1. This x-ray analysis provides the first direct structural evidence of thiazoline ring chemistry in an intein (Fig. 3A), and accounts for the observed 18 Da mass discrepancy (Fig. 2). As described earlier, a thiazoline ring would be expected to form at this position due to aberrant dehydration of the tetrahedral intermediate that follows the initial nucleophilic attack in the first step of splicing (Fig. 1B and Fig. 1C(ii) to (v) to (vi)). An O-protonation event rather than N-protonation is expected to result in the collapse of the tetrahedral intermediate to a thiazoline ring (Fig. 1C(vi)), rather than the resolution to a thioester typical of the splicing pathway (Fig 1C(ii) to (iii) to (iv)). The dehydration reaction accounts for the loss of 18 Da by the thiazoline relative to the thioester forms of the intein. As the thiazoline ring can only be accessed via the tetrahedral intermediate, the aberrant formation of the heterocycle adds to the preponderance of evidence for the existence of the tetrahedral intermediate in the splicing pathway (Fig. 1C(ii)).

The unambiguous high-resolution structure of the thiazoline ring in the N-terminal region indicates a population of homogeneous ring-containing protein (Fig. 3A, Table S1). This result suggests that under the purification and crystallization conditions (pH 8.0), the thiazoline ring is thermodynamically favored and does not undergo interconversion between the thiazoline, the oxythiazolidine intermediate, and the free thiol.

Interestingly, an analogous B-block mutation (T69A) was studied by crystallography in the *Ssp* DnaE intein, yet that crystal structure shows no evidence of thiazoline ring formation (Callahan et al. 2011). Instead, a strain on the scissile N-terminal junction bond was demonstrated to be relaxed when the B-block Thr was mutated to Ala (Fig. S1) (Dearden et al. 2013), suggesting that the role of the B-block Thr in the first step of splicing may vary between inteins. This Thr likely plays an important role in priming the scissile bond for nucleophilic attack in the *Ssp* DnaE intein, whereas it serves to coordinate the resolution of the tetrahedral intermediate as a thioester in the *Mtu* RecA intein.

Several other studies have included Thr-to-Ala mutations at this B block position and invariably report splicing and/or cleavage incompetence (Pearl et al. 2007; Schwarzer et al. 2012). However, these phenotypes were determined by SDS-PAGE rather than via crystallography or mass spectrometry and it is therefore unclear whether the lack of N-S acyl shift is due to prevention of the initial nucleophilic attack or due to formation of a thiazoline ring. Thus, it is presently unclear how prevalent catalytic strategies similar to RecA, which utilize the B-block Thr to direct resolution of the tetrahedral intermediate rather than to destabilize the -1 scissile bond, may be within the intein population as a whole.

The B-block Thr Directs the Fate of the Tetrahedral Intermediate in the RecA Intein

Mutational Analysis at T70 Position: The mutants of the parental I-SM mini-intein construct analyzed here (Fig. 4 and Figs. S3–S9), contain a His tag as part of the N-extein as opposed to those discussed above Fig. 2, and are correspondingly larger. We first compared the wild-type T70 construct containing the N139 mutation and the T70A mutant (without the N139 mutation) with the T70A/N139A I-TM double mutant (Fig. 4, top 3 lines; Fig. S3–S5, respectively). As previously noted, the T70 control with only the N139A mutation appeared as the N-terminal cleavage product (Fig. 4/Fig. S3, $m_{\text{obs}} = 15,877.21$, $m_{\text{calc}} = 15,877.35$ Da, monoisotopic mass). In contrast, the T70A mutant (with wild-type N139) results in a thiazoline on the C-terminal cleavage product (Fig. 4/Fig. S4, $m_{\text{obs}} = 17,714.18$ Da; $m_{\text{calc}} = 17,714.23$ Da, monoisotopic mass). C-terminal cleavage is relieved by the N139 mutation, yielding a thiazoline-containing product, 18 Daltons less than predicted (Fig. 4/Fig. S5, $m_{\text{obs}} = 18,204.39$ Da; $m_{\text{calc}} = 18,204.48$ Da, monoisotopic mass). As shown in Fig. 3B and C, the mutation of the B-block Thr to an Ala (in the absence and presence of mutation N139A) changes both the steric contribution of the sidechain as well as its ability to form H-bonds

To resolve if one of these changes was the determining factor in redirection of the intermediate to a thiazoline ring, three additional mutants, T70V, T70C, and T70S, were generated and tested for the presence of thiazoline ring. The Val mutant sidechain is expected to have an approximately equal steric contribution as Thr but lacks any H-bonding

potential. The T70C mutation alters the steric contribution compared with the Thr sidechain in terms of sidechain sterics but also lacks H-bonding properties. The T70S mutation alters the steric contribution similarly to Cys but maintains the H-bonding properties of the sidechain hydroxyl. Each mutant was overexpressed in *E. coli*, purified at pH 8, and analyzed via mass spectrometry.

The T70V/N139A mutation, like T70A/N139A, resulted in a homogenous intein precursor population that exhibited the -18 Da aberrant mass characteristic of thiazoline ring formation (Fig. 4/ Fig. S6, $m_{\text{obs}} = 18,230.46$ Da, $m_{\text{calc}} = 18,230.51$ Da monoisotopic mass). Thus, it is unlikely that purely steric effects of the sidechain are responsible for preventing the aberrant chemistry. The T70S/N139A mutant, on the other hand, sprayed as a mixture of two species: thiazoline ring-containing precursor and N-terminal cleavage product (Fig. 4/ Fig. S7, precursor: $m_{\text{obs}} = 18,218.40$ Da; $m_{\text{calc}} = 18,218.34$ Da; and N-terminal cleavage: $m_{\text{obs}} = 15,863.10$ Da; $m_{\text{calc}} = 15,863.33$ Da; monoisotopic mass), while the T70C/N139A mutant exhibited only thiazoline ring containing precursor (Fig. 4/ Fig. S8, $m_{\text{obs}} = 18,234.37$ Da; $m_{\text{calc}} = 18,234.45$ Da; monoisotopic mass). The partial rescue with the Ser mutation but not the Cys (Fig. 4, Figs. S7 and S8) suggests that sidechain H-bonding is necessary, but not sufficient, as precise spatial orientation is required for fully competent intein splicing.

Thiazoline Ring Formation is Independent of Identity of -1 Residue: The sole previously-reported case of aberrant thiazoline formation in an intein was found to be due to the mutation of a native Gly to an Ala at the -1 position of the *Ssp* DnaB intein (Ludwig et al. 2008). This effect was traced to the additional methyl group of the Ala sidechain, as it was predicted to sterically clash with an active-site residue and analogs with methyl groups in alternate locations showed no evidence of thiazoline ring formation. The *Mtu* RecA intein, and therefore also the T70A mutant, contains a Lys at the -1 position. Although this bulky residue is native to the extein, we wondered if the truncation of the extein (to five residues) may cause an artefactual conformation of Lys-1 which induces steric hindrance, as the Ala-1 does in *Ssp* DnaB. To determine if the sidechain of the -1 residue was playing a role in thiazoline ring formation we mutated the Lys at the -1 position to Gly. This mutant was overexpressed in *E. coli*, purified at pH 8, and analyzed via mass spectrometry. Despite the presumed reduction in steric hindrance caused by the mutation, the protein exhibited the -18 Da mass indicative of thiazoline ring (Fig. 4/ Fig. S9, $m_{\text{obs}} = 18,131.34$ Da; $m_{\text{calc}} = 18,131.40$ Da; monoisotopic mass).

This finding is significant for two reasons. First, it suggests that while the *Ssp* DnaE intein, with a native Gly-1, is sensitive to steric changes at that position, the catalytic mechanism of the *Mtu* RecA intein, with a bulkier native Lys, has evolved in such a way that the active site is not subject to the same steric constraints. Second, it indicates that in the case of the RecA intein, thiazoline ring formation is caused solely by the T70A mutation and is decoupled from the identity of the -1 extein residue. This independence is highly desirable if an intein is to be used as a modular tool for biotechnological applications.

Crystal Structure Shows Active Site H-bonding Network is

Disrupted: Comparing the crystal structure of the thiazoline ring-containing (T70A) RecA intein with the structure of native intein (T70) (Van Roey et al. 2007) reveals several

Author Manuscript

differences in hydrogen bonding at the N-terminal splice junction (Fig. 3B,C). Somewhat surprisingly, the position of the highly conserved His73 is almost unperturbed with the mutation. It was suspected that a steric clash involving this residue was responsible for thiazoline ring formation in the *Ssp* DnaB intein (Schwarzer et al. 2012), which does not appear to be the case with the RecA intein. Instead, the major changes occur in the hydrogen bonding network. In the native intein, the B-block Thr side chain hydroxyl is in a position to hydrogen bond to both the carbonyl oxygen and amide nitrogen of the Cys1 backbone (Fig. 3B). This potential interaction is absent in the T70A mutant and the Cys1 backbone amide is instead hydrogen bonded to a water molecule, which is not present in the native crystal structure (Fig. 3C). The loss of these two hydrogen bonds likely results in sufficient perturbation of the orientation of the tetrahedral intermediate such that the amide nitrogen is no longer positioned correctly for protonation or the oxygen anion is more favorably positioned for protonation, resulting in the aberrant thiazoline chemistry.

Author Manuscript

The loss of hydrogen bonding may also compromise the compactness of the active site, allowing access to several waters, which are excluded from the native structure, any of which may play a role in the spurious O-protonation event. Additional peripheral changes include a rearrangement of the sidechains of Asp72 and Asp121 (Asp121 corresponds to Asp422 in the full-length RecA intein (Van Roey et al. 2007; Wood et al. 1999)), which have been implicated in catalysis and may play a role in stabilizing the intermediate. An additional change is a rearrangement in the sidechain of Arg53 (Fig 3D). This Arg residue is not part of a conserved block and has therefore received little attention, yet it is in an appropriate position to participate as part of an oxyanion hole. Several inteins, including the *Ssp* DnaE intein, have residues with positively charged sidechains in similar positions (Fig. S1). Each of the aforementioned residues, along with His73, may assist, directly or indirectly, with O- or N-protonation of the anion intermediate. These residues constitute prime candidates for mutation to determine their potential contributions to thiazoline ring formation, and more generally, their atomistic roles in the first step of splicing.

Author Manuscript

It is clear that T70 directs the fate of the tetrahedral intermediate. Based on the mutational studies and the crystal structure, we propose that this effect is mediated through hydrogen bonding, which serves to orient Cys1 directly (and by extension—the tetrahedral intermediate itself), to orient a supporting cast of catalytic residues in relation to the intermediate, and to compact the active site to exclude spurious waters.

B-block Thr is Not Conserved in Non-canonical Splicers: The B-block threonine is highly conserved, occurring in about 80% of inteins (Novikova et al. 2016; Perler 2002) (Fig. 5A). In inteins where the Thr is not conserved, the next two most frequent replacements are Ser and Asp, both of which have side chains capable of hydrogen bonding. Inteins can be divided into two major categories: canonical splicers, which initiate splicing by nucleophilic attack of Cys1, and non-canonical splicers, which initiate splicing by nucleophilic attack of a residue in an alternative position.

Author Manuscript

Separating the populations of canonical and non-canonical splicers reveals a striking difference in B-block conservation—the highly conserved Thr in canonical inteins is replaced by a less conserved Asp in non-canonical inteins (42%) (Fig. 5B). This observation

can be interpreted in a manner consistent with our proposed role for the B-block Thr. Due to differences in positioning of the attacking nucleophiles between canonical and non-canonical inteins, one would expect the orientation of the resulting tetrahedral intermediates to also differ. While intermediates resulting from attack by the 1 position may be well-suited for alignment by the T70 H-bonding network, the intermediates resulting from attacks from other angles are likely to require alternate alignments. In these cases, it is reasonable that the more highly conserved Asp is favorable, or that stabilization is achieved in a different manner. In either case, there is no longer selective pressure to maintain a Thr in the B-block position and conservation is not expected. However, since initial destabilization of the scissile peptide bond prior to nucleophilic attack is expected to be equally important for both canonical and non-canonical inteins, one would expect conservation of residues involved in this mechanism (Fig. S10).

Thiazoline Ring Is Hydrolyzed Under Acidic Conditions

The primary challenge facing intein application is spacio-temporal control of the typically autocatalytic splicing or cleavage activity. Some demonstrated strategies have included modulation of split intein association, reversible inhibition by metal ions, ligand induced structural rearrangements, salt induction and direct chemical modification of catalytic residues (Buskirk et al. 2004; Ciragan et al. 2016; Lew and Paulus 2002; Shi et al. 2014; Skretas and Wood 2005; Wong et al. 2015; Wood and Camarero 2014). For applications involving N-terminal cleavage this chemical modification is typically performed on the attacking nucleophile at the first position of the intein. Examples include the replacement of the native cysteine with a protected light-sensitive derivative or installation of a local redox trap that captures the cysteine in a disulfide (Bocker et al. 2015; Callahan et al. 2013; Ren et al. 2015).

In many ways, the propensity of the T70A intein to form a thiazoline ring can be considered a similar precursor trap. However, the aforementioned modifications are practically useful because they can be reversed by irradiation with light or a reducing environment. If the thiazoline mutant were to be similarly useful, it would be required that upon a change in conditions, first, the ring can be hydrolyzed to tetrahedral intermediate or free thiol and second, the tetrahedral intermediate can subsequently be resolved into thioester rather than returning to thiazoline ring (Fig. 1(vi) to (ii) to (iv)). Thus, we were interested in the stability and reversibility of the chemistry since it does not require specific extein residues or a genetic system capable of utilizing non-native amino acids.

It is well established that thiazoline rings can be opened by acid catalyzed hydrolysis (Gaumont et al. 2009; Ishihara and Shimura 1988). The I-TM T70A/N139A mutant was purified away from the His tag (*cf.* Fig. 2) in phosphate buffer at pH 8, then buffer exchanged into ammonium acetate and the pH adjusted with trifluoroacetic acid. The protein samples were incubated at 4°C and the ratio of ring-containing to ring-hydrolyzed species was periodically determined via mass spectrometry. The thiazoline ring was highly stable in the acetate buffer at pH 7 (Fig. 6A(i) and 6B(i)), but was rapidly hydrolyzed at pH 2 (Fig. 6A(ii) and (iii)). Only minimal ring hydrolysis (<10%) was observed over 7 days at pH 7. At pH 4, hydrolysis occurs but at a very limited rate with a half-life for the reaction of

approximately 4.8 days (Fig. 6B(ii)). At pH 2.5, hydrolysis occurs rapidly with a half-life of 6.5 hours (Fig. 6B(iii)). The large differences in half-life of reaction indicate that the intein can be effectively thiazoline-trapped at relevant protein expression pHs and durations, then switched to ring-hydrolyzed state at low pH.

In addition to the kinetics, the effect of pH on thermodynamics was also investigated at 4°C (Fig. 6C). Thiazoline ring was favored at pH 7. Between pH 6 and pH 4 there was a transition that favors hydrolysis, although about 30% of the population retained the ring. At pH 4 another transition occurred, and the hydrolyzed species was solely favored.

Taken together, the kinetics and thermodynamics describe three major regimes of the reaction. The first regime occurs at ~pH 7 and above, where the rate of hydrolysis is negligible, resulting in a nearly homogeneous ring-containing population. The second regime occurs between pH 6 and pH 4. Here, hydrolysis is favored, but occurs very slowly. The final regime occurs below pH 4 where hydrolysis is favored and occurs rapidly.

The transitions between regimes occur at pH 6 and pH 4, suggesting that His and Asp sidechains may play a direct catalytic role, especially in the case of Asp with the dramatic rate change below pH 4. This observation bolsters the case for examining the two Asp residues that interact with the active site, Asp72 and Asp121 (Fig. 3B–D). The potential role of Asp in cyclization draws an interesting parallel to nonribosomal peptide synthetases, where it has recently been shown that an Asp residue is essential for heterocyclization (Bloudoff and Schmeing 2017; Gehring et al. 1998b). Reorientation of an intein Asp due to the T70A mutation may allow it to play the same role in the intein and suggests the possibility of repurposing active sites for alternate chemistry.

Cleavage Competence of Thiazoline Mutants

Although the T70A mutant exhibited ring hydrolysis at low pH, attempts to induce cleavage with an external nucleophile were unsuccessful, indicating that this mutant is incapable of proceeding to the thioester step of intein chemistry (Fig. 1). We suspected that alternate amino acid substitutions, specifically, those containing H-bonding sidechains, at the Thr B-block position, may form thiazoline ring at high pH, but retain some level of cleavage competence in species in which the ring has been hydrolyzed. Interestingly, when the B-block Thr was replaced by a Ser or Asp, the hydrolyzed species regained moderate catalytic activity at high pH and formed a labile thioester at the N-extein junction. This labile bond could be subsequently cleaved with a nucleophile such as DTT at 4°C (Fig. 7). Increased temperature also reduces ring stabilization through hydrolysis (Fig. S11). Because high pH favors both thiazoline ring formation and cleavage activity, a pH cycling approach is suggested to enhance the yield of cleaved product.

Proposed pH cycling strategy

Under the proposed strategy, an intein precursor is isolated from *E. coli* lysate at neutral to slightly basic pH (pH 7–8) (Fig. 8). This precursor is trapped by a thiazoline ring (induced by the mutation T70A or another T70 mutant) at the N-terminal junction (Fig. 8(i)), which connects the protein of interest (N-extein) to the self-cleaving affinity tag (intein-C-extein). The pH is subsequently lowered (potentially as low as pH 2), resulting in thiazoline ring

hydrolysis (Fig. 8(ii)). Cleavage does not proceed at the low pH, presumably due to insufficient nucleophilicity of Cys1 to start the reaction. To initiate cleavage, the solution must be returned to neutral pH with a nucleophilic agent, e.g. DTT or hydroxylamine, at which point the competing reactions of ring reformation (a reversible process) and cleavage (an irreversible process) proceed (Fig. 1(C)(vi) to (ii) to (iv)). The portion of the population that has undergone cleavage, generating free protein of interest (Fig. 8(iii)), may be collected, while the precursor population in which the ring has reformed (Fig. 8(i)) may be fed back into the pH cycle for another round of ring hydrolysis followed by cleavage/ring reformation. The process may be repeated as many times as necessary to yield a sufficient amount of cleaved product (Fig. 8(iii)). While the low pH range is a benefit for this system, it also serves as a general limitation, preventing application to proteins of interest that are susceptible to denaturation at low pH. However, due to the robust trapping of precursor, this strategy may be useful for proteins tolerant to low pH, particularly when premature cleavage is undesirable, such as the expression of toxic proteins. The strategy may also be useful in systems that achieve low pH over the course of expression, and may be tuned to more favorable pH ranges by targeted mutagenesis.

CONCLUSIONS

We have engineered an intein that autocatalytically forms a thiazoline ring at the N-terminal splicing junction. The single position responsible for the aberrant chemistry, a highly conserved B-block Thr, has been shown to “spring-load” the scissile N-terminal peptide bond in another intein. Here, the primary role is to correctly direct the tetrahedral intermediate formed during the first step of splicing to a thioester that induces cleavage and away from off-pathway collapse to a thiazoline ring. Structural and mutational studies suggest that this Thr, when present, mediates correct alignment of the active site through hydrogen bonding interactions. The thiazoline ring was visualized crystallographically at 1.54 Å resolution. We also demonstrated the stability of the thiazoline ring at neutral pH, as well as hydrolytic opening of the ring and subsequent thioester formation under acidic and increased temperature conditions. Such properties could be harnessed for use in biotechnological applications requiring an activity switch for cleavage.

MATERIALS AND METHODS

Materials and Bacterial Strains

Oligonucleotide primers were synthesized by IDT. Restriction enzymes and ligase were obtained from New England Biolabs. MasterMix with Taq polymerase (ThermoSci) was used for PCR. The QuikChange Lightning Single-site Kit (Agilent) was used for site-directed mutagenesis. The vector pET30b was purchased from Novagen. HisTrapHP columns (GE Healthcare) and Amicon Ultra Centrifugal Filter units (EMD Millipore) were used for protein purification and concentration. ZebaSpin columns (ThermoSci) were used for protein desalting. Mass spectrometry grade ammonium acetate and formic acid were purchased from Sigma. *E. coli* DH5 α was used for cloning and *E. coli* BL21(DE3) was used for protein expression. All plasmids were verified by DNA sequencing (Eton Biosciences).

Construction of I-SM and I-TM Intein Mutants

The intein used in this work was based on a splicing competent minimized RecA intein I-SM, which has been extensively characterized (Hiraga et al. 2005; Van Roey et al. 2007). The I-SM intein was PCR amplified with a forward primer containing an NdeI site, a 6xHis tag, a TEV cleavage site, and 5 native N-extein residues and a reverse primer containing a HindIII site, 5 native C-extein residues and an N139A mutation. Additionally, a T70A mutation was introduced by SOEing PCR, generating the I-TM construct. The sequence of this construct, with intein residues in upper case, is as follows: mhhhhhhenlyfqgvvknkCLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVA^AAKDGTLHARPVVSWFDQGT^RRDVIGLRIAGGAILWA^APDHKVLTEYGWRAAGELRKGD^RVAVRDVETGELRYSVIREVLP^TRRARTFDLEVEELHTLVAEGVVVH^Acsp^pf. The T70A and N139A mutations of I-TM are underscored. This PCR to generate I-TM was digested with NdeI and HindIII, gel purified, then ligated into the ORF of the pET30b vector which had been similarly treated. Other T70 mutants, including T70D and T70S, were made by following the standard Agilent QuikChange Lightning protocol for site-directed mutagenesis using primers designed with the Agilent Online tool. A similar procedure was followed to generate the C1A/C+1A mutant using the multisite-directed mutagenesis kit and online primer design tool.

Expression and Purification of I-TM intein

The I-TM containing plasmid was transformed into *E. coli* BL21(DE3) cells which were grown to an OD₆₀₀ of 0.5 at 37°C. The cells were then induced with 0.4 mM IPTG. Induction was carried out at 16°C overnight and the cells were pelleted via centrifugation and stored at -80°C. The cells were lysed via sonication and clarified lysate was loaded on a pre-equilibrated NiNTA column. The collected elution fractions were assayed for purity and acceptable fractions were treated overnight with TEV protease at 4°C. These fractions were lowered to 75 mM imidazole by dilution then run over another NiNTA column. Flowthrough, containing purified affinity-tagless I-TM was collected and concentrated in 3 kDa cutoff Amicon centrifugal filter units.

Mass Spectrometry

The affinity-purified I-TM intein was treated with 1 mM EDTA and 1 mM TCEP for one hour, then buffer exchanged into 150 mM ammonium acetate using a ZebaSpin desalting column. The protein was further desalted by five rounds of 5x concentration and dilution in 0.5 mL Amicon Ultra concentrators. If pH modulation of the ammonium acetate buffer was necessary, it was adjusted using formic acid or ammonium hydroxide. The protein was analyzed on an LTQ Orbitrap Velos instrument (Thermo Scientific, Grand Island, NY) in the positive ion mode. All analyses were performed in nanoflow electrospray ionization (ESI) mode by using quartz emitters produced in house. Up to 5 nL samples were typically loaded into each emitter by using a gel-loader pipette tip. A stainless-steel wire was inserted in the back-end of the emitter to supply an ionizing voltage that ranged between 0.9 and 1.2 kV. Protein masses were calculated based on average mass.

Crystallization, structure determination and refinement

The affinity-purified I-TM intein was further purified by a gel filtration 16/60 Superdex 75 column (GE HealthCare). The fractions containing intein were pooled together and concentrated to 13 mg/mL in a buffer composed of 20 mM Tris, pH 8.0, 100 mM NaCl, 2 mM DTT. Initial crystallization conditions were obtained by screening the Hampton Research Index HT crystallization screen and existing crystallization conditions for RecA intein variants (Van Roey et al. 2007), using the hanging-drop vapor diffusion method. Upon optimization, large crystals were grown by mixing 1 μ L of intein and 1 μ L of reservoir solution containing 44% (w/v) ammonium sulfate, 0.1 M Hepes, pH 7.5, 10% glycerol. The crystals of intein belong to space group $P2_12_12_1$, with unit cell parameters of $a=56.65$ Å, $b=64.63$ Å, $c=37.17$ Å, $\alpha=\beta=\gamma=90^\circ$ and one molecule per asymmetric unit cell. Prior to data collection, all crystals were transferred to a cryo-protectant solution containing crystallization buffer supplemented with 25% glycerol. The crystals were flash-cooled directly in liquid nitrogen. Diffraction data for the intein crystals were collected at 1.54 Å resolution at 100K at beamline X25 of the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL). Data were processed, scaled, and reduced using the programs HKL2000 (Otwinowski and Minor 1997) and Phenix suite (Adams et al. 2010). The structure of the intein was determined using molecular replacement method with PDB 2IN0 as a search model using the PHASER program suite (McCoy et al. 2005). Structure refinement was carried out using the Phenix program suite (Table S1). At 1.54 Å resolution, the final R_{cryst} was 19.6%, with an R_{free} of 21.3%.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institute of Health (GM44844 to M.B and NIGMS, R01 GM121844–01 and NIAID, R21 AI133617–01 to D.F.). Brian Callahan's seminal ideas are appreciated. C.S.P. was supported by the Howard P. Isermann Fellowship and Biomolecular Science and Engineering Training Grant GM067545–10. Funds from G.B.'s Institute Chair at RPI are also acknowledged. We thank the staff members at the beamline X25 at the National Synchrotron Light Source for help in data collection. The National Synchrotron Light Source Beamline X25 is supported by the United States Dept. of Energy, Office of Biological and Environmental Research, the National Institute of General Medical Sciences, and the National Institutes of Health National Center for Research Resources.

REFERENCES

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW and others. 2010 PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2), 213–21. [PubMed: 20124702]
- Banki MR, Gerngross TU, Wood DW. 2005 Novel and economical purification of recombinant proteins: Intein-mediated protein purification using in vivo polyhydroxybutyrate (PHB) matrix association. *Protein Science*, 14, 1387–1395. [PubMed: 15883185]
- Bloudoff K, Schmeing TM. 2017 Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim Biophys Acta*, 1865, 1587–1604.

- Bocker JK, Friedel K, Matern JC, Bachmann AL, Mootz HD. 2015 Generation of a genetically encoded, photoactivatable intein for the controlled production of cyclic peptides. *Angew Chem Int Ed Engl*, 54, 2116–20. [PubMed: 25557882]
- Buskirk AR, Ong Y-C, Gartner ZJ, Liu DR. 2004 Directed evolution of ligand dependence: Small-molecule-activated protein splicing. *PNAS*, 101, 6. [PubMed: 14695894]
- Calderone TL, Stevens RD, Oas TG. 1996 High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. *J Mol Biol*, 262, 407–12. [PubMed: 8893852]
- Callahan BP, Stanger M, Belfort M. 2013 A redox trap to augment the intein toolbox. *Biotechnol Bioeng*, 110, 1565–73. [PubMed: 23280506]
- Callahan BP, Topilina NI, Stanger MJ, Van Roey P, Belfort M. 2011 Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications. *Nat Struct Mol Biol*, 18, 630–3. [PubMed: 21460844]
- Chan H, Pearson CS, Green CM, Li Z, Zhang J, Belfort G, Shekhtman A, Li H, Belfort M. 2016 Exploring intein inhibition by platinum compounds as an antimicrobial strategy. *J Biol Chem*, 291, 22661–22670. [PubMed: 27609519]
- Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA and others. 1997 Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, 192, 271–81. [PubMed: 9224900]
- Ciragan A, Aranko AS, Tascon I, Iwai H. 2016 Salt-inducible protein splicing in cis and trans by inteins from extremely halophilic archaea as a novel protein-engineering tool. *J Mol Biol*, 428, 4573–4588. [PubMed: 27720988]
- Dearden AK, Callahan B, Roey PV, Li Z, Kumar U, Belfort M, Nayak SK. 2013 A conserved threonine spring-loads precursor for intein splicing. *Protein Sci*, 22, 557–63. [PubMed: 23423655]
- Evans TC, Jr., Benner J, Xu MQ. 1998 Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci*, 7, 2256–64. [PubMed: 9827992]
- Gaumont AC, Gulea M, Levillain J. 2009 Overview of the chemistry of 2-thiazolines. *Chem Rev*, 109, 1371–401. [PubMed: 19154153]
- Ge X, Yang DSC, Trabbic-Carlson K, Kim B, Chilkoti A, Filipe CDM. 2005 Self-cleavable stimulus responsive tags for protein purification without chromatography. *J Amer Chem Soc*, 127, 11228–11229. [PubMed: 16089436]
- Gehring AM, DeMoll E, Fetherston JD, Mori I, Mayhew GF, Blattner FR, Walsh CT, Perry RD. 1998a Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem Biol*, 5, 573–86. [PubMed: 9818149]
- Gehring AM, Mori II, Perry RD, Walsh CT. 1998b The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry*, 37, 17104. [PubMed: 9836605]
- Hiraga K, Derbyshire V, Dansereau JT, Van Roey P, Belfort M. 2005 Minimization and stabilization of the *Mycobacterium tuberculosis* recA intein. *J Mol Biol*, 354, 916–26. [PubMed: 16288917]
- Hiraga K, Soga I, Dansereau JT, Pereira B, Derbyshire V, Du Z, Wang C, Van Roey P, Belfort G, Belfort M. 2009 Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. *J Mol Biol*, 393, 1106–17. [PubMed: 19744499]
- Hirata R, Ohsumi Y, Nakano A, Kawasaki H, Suzuki K, 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, 6726–33. [PubMed: 2139027]
- Ishihara H, Shimura K. 1988 Further evidence for the presence of a thiazoline ring in the isoleucylcysteine dipeptide intermediate in bacitracin biosynthesis. *FEBS Lett*, 226, 319–23. [PubMed: 3338562]
- Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH. 1990 Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*, 250, 651–7. [PubMed: 2146742]

- Kawasaki M, Nogami S, Satow Y, Ohya Y, Anraku Y. 1997 Identification of three core regions essential for protein splicing of the yeast Vma1 protezyme. A random mutagenesis study of the entire Vma1-derived endonuclease sequence. *J Biol Chem*, 272, 15668–74. [PubMed: 9188457]
- Kelly WL. 2012 Biosynthesis: Ringing in a new view. *Nat Chem Biol*, 8, 505–7. [PubMed: 22596200]
- Klabunde T, Sharma S, Telenti A, Jacobs WR, Jr., Sacchettini JC. 1998 Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat Struct Biol*, 5, 31–6. [PubMed: 9437427]
- Lew BM, Paulus H. 2002 An in vivo screening system against protein splicing useful for the isolation of non-splicing mutants or inhibitors of the RecA intein of *Mycobacterium tuberculosis*. *Gene*, 282, 169–77. [PubMed: 11814689]
- Ludwig C, Schwarzer D, Mootz HD. 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, 25264–72. [PubMed: 18625708]
- McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. 2005 Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr*, 61(Pt 4), 458–64. [PubMed: 15805601]
- Mizutani R, Anraku Y, Satow Y. 2004 Protein splicing of yeast VMA1-derived endonuclease via thiazolidine intermediates. *J Synchrotron Radiat*, 11(Pt 1), 109–12. [PubMed: 14646148]
- Mizutani R, Nogami S, Kawasaki M, Ohya Y, Anraku Y, 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, 919–29. [PubMed: 11884132]
- Muir TW, Sondhi D, Cole PA. 1998 Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A*, 95, 6705–10. [PubMed: 9618476]
- Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI, Belfort M. 2016 Intein clustering suggests functional importance in different domains of life. *Molecular Biology and Evolution*, 33, 783–799. [PubMed: 26609079]
- Otwinowski Z, Minor W. 1997 Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 276, 307–326.
- Paulus H. 2000 Protein splicing and related forms of protein autoprocessing. *Annual Review of Biochemistry*, 69, 447–496.
- Pearl EJ, Tyndall JD, Poulter RT, Wilbanks SM. 2007 Sequence requirements for splicing by the Cne PRP8 intein. *FEBS Lett*, 581, 3000–4. [PubMed: 17544410]
- Pereira B, Shemella PT, Amitai G, Belfort G, Nayak SK, Belfort M. 2011 Spontaneous proton transfer to a conserved intein residue determines on-pathway protein splicing. *J. Mol. Biol*, 406, 13.
- Perler FB. 2002 InBase: the intein database. *Nucleic Acids Res*, 30, 383–4. [PubMed: 11752343]
- Perler FB, Adam E. 2000 Protein splicing and its applications. *Current Opinion in Biotechnology*, 11, 377–383. [PubMed: 10975457]
- Perler FB, Olsen GJ, Adam E. 1997 Compilation and analysis of intein sequences. *Nucleic Acids Res*, 25, 1087–93. [PubMed: 9092614]
- Petrokovski S. 1994 Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci*, 3, 2340–50. [PubMed: 7756989]
- Petrokovski S. 1998 Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci*, 7, 64–71. [PubMed: 9514260]
- Poland BW, Xu MQ, Quiocho FA. 2000 Structural insights into the protein splicing mechanism of PI-SceI. *J Biol Chem*, 275, 16408–13. [PubMed: 10828056]
- Ren W, Ji A, Ai HW. 2015 Light activation of protein splicing with a photocaged fast intein. *J Am Chem Soc*, 137, 2155–8. [PubMed: 25647354]
- Saleh L, Perler FB. 2006 Protein splicing in cis and in trans. *Chemical Record* 6(4):183–193. [PubMed: 16900466]
- Schwarzer D, Ludwig C, Thiel IV, Mootz HD. 2012 Probing intein-catalyzed thioester formation by unnatural amino acid substitutions in the active site. *Biochemistry*, 51, 233–42. [PubMed: 22182201]
- Shah NH, Muir TW. 2014 Inteins: Nature's gift to protein chemists. *Chem Sci*, 5, 446–461. [PubMed: 24634716]

- Shi C, Tarimala A, Meng Q, Wood DW. 2014 A general purification platform for toxic proteins based on intein trans-splicing. *Appl Microbiol Biotechnol*, 98, 9425–35. [PubMed: 25296714]
- Skretas G, Wood DW. 2005 Regulation of protein activity with small-molecule-controlled inteins. *Protein Sci*, 14, 523–32. [PubMed: 15632292]
- Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB. 1999 Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *Biotechniques*, 27, 110–4. [PubMed: 10407673]
- Topilina NI, Mills KV. 2014 Recent advances in in vivo applications of intein-mediated protein splicing. *Mob DNA*, 5, 5. [PubMed: 24490831]
- Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V. 2007 Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *J Mol Biol*, 367, 162–73. [PubMed: 17254599]
- Volkman G, Liu XQ. 2011 Intein lacking conserved C-terminal motif G retains controllable N-cleavage activity. *FEBS J*, 278, 3431–46. [PubMed: 21787376]
- Walsh CT, Malcolmson SJ, Young TS. 2012 Three ring posttranslational circuses: insertion of oxazoles, thiazoles, and pyridines into protein-derived frameworks. *ACS Chem Biol*, 7, 429–42. [PubMed: 22206579]
- Wong S, Mosabbir AA, Truong K. 2015 An engineered split intein for photoactivated protein trans-splicing. *PLoS One*, 10, e0135965. [PubMed: 26317656]
- Wood DW, Camarero JA. 2014 Intein applications: from protein purification and labeling to metabolic control methods. *J Biol Chem*, 289, 14512–9. [PubMed: 24700459]
- Wood DW, Derbyshire V, Wu W, Chartrain M, Belfort M, Belfort G. 2000 Optimized single-step affinity purification with a self-cleaving intein applied to human acidic fibroblast growth factor. *Biotechnol Prog*, 16, 1055–63. [PubMed: 11101334]
- Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M. 1999 A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol*, 17, 889–92. [PubMed: 10471931]
- Xu MQ, Perler FB. 1996 The mechanism of protein splicing and its modulation by mutation. *EMBO J*, 15, 5146–53. [PubMed: 8895558]
- You J, Cohen RE, Pickart CM. 1999 Construct for high-level expression and low misincorporation of lysine for arginine during expression of pET-encoded eukaryotic proteins in *Escherichia coli*. *Biotechniques*, 27, 950–4. [PubMed: 10572642]
- Zhang L, Zheng Y, Callahan B, Belfort M, Liu Y. 2011 Cisplatin inhibits protein splicing, suggesting inteins as therapeutic targets in mycobacteria. *J Biol Chem*, 286, 1277–82. [PubMed: 21059649]

(A) Schematic sequence for the RecA intein with N- and C-exteins. Only the residues with numbers and motif locations of interest are shown. Vertical arrows identify the N- and C-terminal splicing junctions. This construct represents a RecA mini-intein from which the endonuclease has been deleted between block B and F. (B) Overall splicing pathway: The process begins with a nucleophilic attack on the adjacent peptide bond in the precursor (P) by Cys1 of the intein (first step, shaded), which subsequently undergoes a transesterification mediated by Cys+1 of the C-extein, resulting in a branched intermediate (BI). The intein is then excised via cyclization of the terminal Asn. The thioester linking the exteins is resolved to a peptide bond by nucleophilic attack of the C-extein amine. (C) Detailed pathway for the first step of splicing (shaded). This step results in formation of a thioester (iv) from a free thiol (i), as well as a postulated aberrant pathway, which results in formation of a thiazoline ring (vi). The fate of the tetrahedral intermediate (ii, TI) is dependent on protonation of the nitrogen (iii) or oxygen (v) of this intermediate. TE, thioester; TR, thiazoline ring.

Figure 2. Electrospray ionization (ESI) mass spectrum of RecA intein with T70A/N139A mutations.

An 18-Da mass reduction of the T70A/N139A mutant represents dehydration associated with thiazoline ring protonation. For the sake of clarity, only the region containing the 7+ charge state is shown. The most intense signal (arrow) corresponds to thiazoline containing intein (monoisotopic $M_{\text{obs}} = 16,454.60$ Da versus $M_{\text{calc}} = 16,454.72$ Da calculated from the sequence). The expected mass of the T70A/N139A mutant without the thiazoline ring was not detected (M_{calc} of 16,472.73 Da). Additional signals were detected for corresponding T70A forms in which Arg-to-Lys (K-R) substitutions were induced during protein expression, which resulted in a mass decrease of 28 Da. Typical salt adducts were detected for Na^+ , K^+ and Na^+/K^+ (DM of +22, +38, and +60 Da, respectively). It should be noted that the high resolution afforded by these determinations allowed us to resolve the isotopic envelope of each detected species. In this way, it was possible to determine the respective charge state from the mass-over-charge ratios (m/z) spacing between contiguous isotopic signals: $z = 1/D_{m/z}$, where z is the charge state and $D_{m/z}$ is the m/z spacing. Inset shows the structure of the thiazoline ring. This construct had the His tag removed by proteolytic cleavage.

Figure 3. Crystal structure of the I-TM intein with the thiazoline ring.

(A) Electron density map of a thiazoline ring. The ring is formed between the first Cys residue of the intein (C1) and the preceding N-extein Lys residue (K-1) in the I-TM (T70A/N139A) construct. (B) and (C) Comparison of H-bonding in the crystal structures of spliced and thiazoline ring containing intein, respectively, with faint background structures comparing T70 with A70. (B) In the native intein the T70 backbone amide is H-bonded to the Cys1 carbonyl carbon, and the sidechain OH forms H-bonds with both the Cys1 carbonyl carbon and amide nitrogen. (C) In the thiazoline mutant the A70 sidechain H-bond to the Cys1 carbonyl oxygen is absent and, interestingly, the H-bond to the amide nitrogen is replaced by a water, which is not present in the thiazoline-less crystal structure. (D) Location of critical residues in the vicinity of the N-terminus of the intein. Asp121 corresponds to Asp422 in the full-length RecA intein (Van Roey et al. 2007; Wood et al. 1999).

Mutant	N-EXTEIN			INTEIN				C-EXTEIN		Calculated monoisotopic Mass	Observed monoisotopic Mass	Error Ppm	Spectra Supplementary Fig. #s
	Residue # -1	1		70				139	+1				
	K	C		T	H		D	N	C				
N139A *										15,877.35	15,877.21	12.6	S3
T70A ^Ψ		⊙								17,714.23	17,714.18	4.7	S4
T70A/N139A		⊙								18,204.48	18,204.39	4.7	S5
T70V/N139A		⊙								18,230.51	18,230.46	3.0	S6
T70S/N139A*		⊙								18,218.34	18,218.40	3.3	S7
										15,863.33	15,863.10	16.6	
T70C/N139A		⊙								18,234.45	18,234.37	4.1	S8
K-1G/T70A/N139A		G⊙								18,131.40	18,131.34	3.5	S9

⊙ = thiazoline ring * = N-terminal cleavage ^Ψ = C-terminal cleavage

Figure 4. Characteristics of RecA mini-intein mutants investigated by ESI-MS.

Full mass range spectra on which masses are based are shown in the Supplementary Figs. S3–S9. From the top, T70 wild type with Thr at position 70 prevents thiazoline ring formation and restores N-terminal cleavage competence (Fig. S3). Wild type Asn at the ultimate 139 position restores C-terminal cleavage but does not prevent thiazoline ring formation (Fig. S4). The T70A, T70V, T70C, and K-1G mutants with N139A all spray as solely thiazoline ring-containing species (Figs. S5, S6, S8 and S9, respectively). Interestingly, the T70S spectra contain both thiazoline ring-containing precursor and N-terminal cleavage product, suggesting partial rescue of the thiazoline phenotype (Fig. S7). The His tag (part of the N-extein) was an integral component of the constructs, explaining the increased masses reported over those in Figures 2 and 6 (see Materials and Methods for sequence).

Figure 5. Conservation of residues in the TxxH motif.

(A) Canonical splicing inteins. Thr is highly conserved in inteins that splice by the canonical mechanism, replaced occasionally by Ser or Asn. These alternate residues, like Thr, are able to form hydrogen bonds which are expected to play an important role in orientation of the tetrahedral intermediate. (B) Non-canonical splicing inteins. Thr conservation is significantly reduced in inteins that splice by alternate mechanisms. Source references (Novikova et al. 2016; Perler 2002).

Figure 6. Stability of thiazoline ring as a function of pH and time.

(A) Thiazoline ring stability by mass spectrometry analysis. Red bars denote thiazoline ring-containing species and blue bars denote hydrolyzed species. The ring is highly stable at 4°C and at pH 7 (i) but hydrolyzes at acidic pH 2 for 15 min (ii), and at pH 2 for 3 h (iii).

Calculated monoisotopic mass of thiazoline-ring containing protein: 16,454.72 Da; the calculated monoisotopic mass of hydrolyzed protein: 16,472.73 Da. (B) Kinetic profiles of ring hydrolysis at pH 7 (i), 4 (ii), and 2.5 (iii) for 7 days. (C) Fraction of thiazoline ring after 6.5 h as a function of solution pH. This construct had the His tag removed by proteolytic cleavage as in Fig. 2.

Figure 7. Behavior of two hydrogen bond-forming mutants in the TxxH motif.

Hydrogen bond-forming mutants (T70S/N139A and T70D/N139A) with the His tag as part of the N-extein, exhibit a propensity to form thiazoline ring but retain the ability to form a thioester when thiazoline ring is hydrolyzed at 4°C. The labile thioester formation is identified by DTT-induced cleavage. Lack of cleavage by TCEP, a non-nucleophilic reducing agent, clarifies that trapping is thiazoline based and not redox based. Cleavage incompetent (C1A) and fully N-terminal cleavage competent (native T70/N139A) inteins serve as controls at 4°C. BI = branch intermediate, P = precursor, C = cleavage product.

Figure 8. Proposed separation protocol: pH cycling to produce a desired product.

We propose that the T70A mutant is used to prevent N-terminal cleavage via the formation of a thiazoline ring. A fusion construct (i) comprising a desired protein (N-ext), a modified mini-intein (int) and a binding domain (C-ext, shown in Fig. 1 but omitted here) for later purification, is over-produced in a bioreactor at pH 7–8. The pH of the solution is then reduced to 2 to hydrolyze the thiazoline ring (ii) and then increased to 7.4 with addition of a nucleophile (e.g. DTT, hydroxylamine etc) to induce irreversible cleavage to obtain the desired product (iii). For those molecules that did not cleave but formed thiazoline ring structures (i), the solution is again dropped to pH 2 to re-enter the cycle. The product-containing solution with the free N-ext plus the int/C-ext fusion (iii) is then passed through an affinity column or membrane to remove the fusion and purify the product of interest.