BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

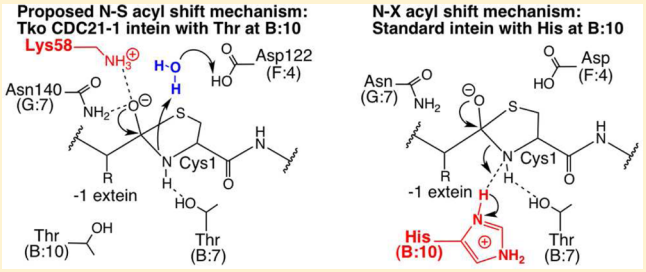
The *Thermococcus kodakaraensis* Tko CDC21-1 Intein Activates Its N-Terminal Splice Junction in the Absence of a Conserved Histidine by a Compensatory Mechanism

Kazuo Tori, Manoj Cheriyan, Chandra Sekhar Pedamallu,[†] Marleny A. Contreras,[‡] and Francine B. Perler*

New England BioLabs, Ipswich, Massachusetts 01938, United States

Supporting Information

ABSTRACT: Inteins and other self-catalytic enzymes, such as glycosylasparaginases and hedgehog precursors, initiate autocleavage by converting a peptide bond to a (thio)ester bond when Ser, Thr, or Cys undergoes an N−[S/O] acyl migration assisted by residues within the precursor. Previous studies have shown that a His at position 10 in intein Block B is essential for this initial acyl migration and N-terminal splice junction cleavage. This His is present in all inteins identified to date except the *Thermococcus kodakaraensis* Tko CDC21-1 intein orthologs and the inactive *Arthrobacter* species FB24 Arth_1007 intein. This study demonstrates that the Tko CDC21-1 intein is fully active and has replaced the lost catalytic function normally provided by the Block B His using a compensatory mechanism involving a conserved ortholog-specific basic residue (Lys_{58}) present outside the standard intein conserved motifs. We propose that Lys_{58} catalyzes the initial N−S acyl migration by stabilizing the thiazolidine−tetrahedral intermediate, allowing it to be resolved by water-mediated hydrolysis rather than by protonating the leaving group as His is theorized to do in many other inteins. Autoprocessing enzymes may have more flexibility in evolving catalytic variations because high reaction rates are not required when performing single-turnover reactions on "substrates" that are covalently attached to the enzyme. Consequently, inteins have more flexibility to sample catalytic mechanisms, providing insight into various strategies that enzymes use to accomplish catalysis.



Intein genes are selfish DNAs inserted by lateral transmission into host genes immediately before a Ser, Thr, or Cys codon.[1−4] The effects of this insertion are minimized because the intein can splice itself out of the host protein by a self-catalytic, post-translational process that generates an active host protein after ligation of the host protein fragments (termed exteins). The reaction is catalyzed without exogenous energy sources or cofactors by the intein together with the C-extein Ser, Thr, or Cys that is always immediately adjacent to the intein (termed the +1 residue). As of February 2012, more than 550 inteins have been identified in archaea, eubacteria, single-cell eukaryotes, phage, and viruses; a listing of inteins and their properties is available in the InBase database (http://www.neb.com/neb/inteins.html).[4] Intein splicing domains contain four sequence motifs (Table 1, Blocks A, B, F, and G) that form part of the protein splicing active site.[3−5] Although not well-conserved in their entirety, each motif contains specific positions with highly conserved residues or groups of similar residues. To simplify comparison of motif residues between inteins of various lengths, positions within these motifs are denoted using the motif name followed by a colon and the position number within the motif (Table 1).[6]

Although inteins do not perform multiple-turnover reactions like classical enzymes, they use similar catalytic strategies to promote peptide bond cleavage and joining.[2,7,8] The "substrates" of the intein are the flanking extein sequences. Three different nucleophiles perform three temporally coordinated nucleophilic displacement reactions at the two scissile bonds between the intein and the exteins (the N- and C-terminal splice junctions). There is variability in the residues and positions within each intein that facilitate these reactions.

The Class 1 intein-mediated protein splicing pathway (Figure 1)[2,7,8] is initiated when the intein N-terminal Ser_1 or Cys_1 undergoes an N−[S/O] acyl migration to form the linear (thio)ester intermediate (**IIa**). In step 2, the first residue of the C-extein (Cys_{+1}, Ser_{+1}, or Thr_{+1}) cleaves the N-terminal splice junction (thio)ester bond, transferring the N-extein to its side chain while forming the Block G branched intermediate (BI, **III**).[9] Thus, for mechanistic studies, the intein enzyme includes the intein that is spliced out plus the C-terminal splice junction extein residue (+1 aa). The BI is resolved by cyclization of the intein C-terminal Asn, yielding ligated exteins bound by a (thio)ester bond (**V**).[10] A native peptide bond is formed between the exteins after a spontaneous [S/O]−N acyl migration (**VI**), driven by the acyl migration equilibrium in

Received: December 13, 2011
Revised: February 27, 2012
Published: March 1, 2012

   dx.doi.org/10.1021/bi201840k | *Biochemistry* 2012, 51, 2496−2505

Table 1. Sequences of Representative Protein Splicing Domain Motifs[a]

Intein	C	Block A	Block B	Block F	Block G
		1	7 10	4	67 8
Tko CDC21-1	1	CVAPDSIIKTNLG	GLSITVTPE**T**KLLT	YVYDLTV**EGSHSFI	NGFVVHN/T
Hwa MCM-2	1	CVTGETRIHTTDG	GKQLEASVN**T**PVLT	EVFDLTV**PNTQNFI	GGIVTHN/T
Pab CDC21-2	1	CVAPETLIITENG	GKQITVTPE**T**KLLT	HVYDLTV**EGSHSFI	NGFVVHN/T
Pfu CDC21	1	CVAPDSLVVVNDK	GKEITVTPE**T**KLLT	YVYDLTV**EGSHSFI	NGFVVHN/T
Pho CDC21-2	1	CVAPDTLINTDNG	GRSIALTRE**T**KLLT	YVYDLTV**EGSHNFV	NGFIVHN/T
Npu DnaE	1	CLSYETEILTVEY	GSVIRATSD**H**RFLT	NVYDIGV**ERDHNFA	NGFIASN/C
Mxe GyrA	1	CITGDALVALPEG	GLRVTGTAN**H**PLLC	PVYSLRV*DTADHAFI	NGFVSHN/T
Sce VMA	1	CFAKGTNVLMADG	LLKFTCNAT**H**ELVV	DYYGITLSDDSDHQFL	NQVVVHN/C
Mtu RecA	1	CLAEGTRIFDPVT	GAIVWATPD**H**KVLT	RTFDLEV**EELHTLV	EGVVVHN/C
Mja KlbA	2	ALAYDEPIYLSDG	RREITLTHD**H**PVYI	HIYDLTV**EDNHTYI	EGFAVSN/C
MP-Be DnaB	3	PLALNTEVPTPSG	GTEITASAS**H**GWTT	PVKCIGI*DTEDHLFQ	SRILTHN/T

[a]Intein names[4] are listed in the first column, followed by the intein class (C) in the second column. The position number in each motif for residues known to be important for catalysis is indicated above the position. Residues within the motifs are denoted using the motif name followed by a colon and the position number within the motif. Class 1 inteins always begin with Ser_1, Thr_1, or Cys_1. Class 2 and 3 inteins begin with other residues. Class 3 inteins always have $Trp_{B:12}$ and $Cys_{F:4}$. The most highly conserved intein residue ($His_{B:10}$) and $Thr_{B:10}$ in the Tko CDC21-1 intein orthologs (Hwa MCD-2, Pab CDC21-2, Pfu CDC21, and Pho CDC21-1 inteins) are in bold. The C-terminal splice junction is indicated by a slash. The asterisk represents an alignment gap.

the absence of catalytic forces.[11] Steps 1 and 2 can be considered a coordinated double-displacement mechanism, while step 3 is a mechanistically unrelated nucleophilic displacement.[8] In Class 1 inteins, the +1 aa cannot attack an amide bond at the N-terminal splice junction.[9] Several studies show coupling of N-terminal and C-terminal splice junction cleavage (on- or off-pathway), while others do not.[12−16] Heterologous exteins often result in inefficient splicing and increased levels of off-pathway cleavage (Figure 2), possibly because of the loss of correct enzyme−substrate interactions.[17,18]

This scenario has become more complicated as more intein variants have been identified. Inteins are now divided into three classes based on differences in nucleophiles, conserved sequences, and protein splicing mechanisms (Figure 1 and Table 1).[6] Class 2 and Class 3 inteins cannot follow the Class 1 pathway because they have residues at their N-termini that cannot form the linear (thio)ester intermediate (**IIa**) (Figure 1). In Class 2 inteins, such as the *Methanococcus jannaschii* Mja KlbA intein, the +1 nucleophile directly attacks an amide bond at the N-terminal splice junction (**Ib**).[19] Class 3 inteins, such as the mycobacteriophage Bethlehem MP-Be DnaB intein, cleave the N-terminal splice junction using Cys at position 4 of Block F (F:4) to form the Block F BI (**IIc**).[6,20,21]

Many proteins initiate autoprocessing reactions by the same N−[S/O] acyl migration as Class 1 inteins, including glycosylasparaginases, pyruvoyl enzymes, α-macroglobulins, nucleoporin Nup145p, and the Hedgehog signaling proteins.[2,7,8,22−25] Nature has repeatedly used this strategy to cleave peptide bonds in the absence of an exogenous enzyme. Mutagenesis, structural, thermodynamic, and kinetic approaches have all been used to examine the residues within the intein that assist this seminal acyl migration. Intein structures[18,26−33] and mutagenesis studies[6,19−21,30,33,34] indicate that the His at Block B position 10 ($His_{B:10}$) plays an essential role in activating the N-terminal splice junction, but its precise mechanistic role is debated and may vary among inteins. The residue at Block B position 7 (B:7) is not essential, and the effects of mutating it vary in different inteins.[2,7,8,23] The most common catalytic role attributed to $His_{B:10}$ is protonating the amide nitrogen leaving group of the tetrahedral intermediate in the linear (thio)ester intermediate (**IIa**).[18,26,27,29−32]

$His_{B:10}$ is present in all inteins currently in InBase except for the *Arthrobacter* species Arsp-FB24 Arth_1007 intein and the *Thermococcus kodakaraensis* KOD1 Tko CDC21-1 intein orthologs.[4] The Arsp-FB24 Arth_1007 intein is an inactive pseudogene that not only lacks $His_{B:10}$ but also is missing other essential catalytic residues.[36] In contrast, the Tko CDC21-1 intein orthologs retain all of the other known essential intein residues (Table 1). This study examines whether the Tko CDC21-1 intein is active despite having Thr_{81} at Block B position 10 instead of His.

EXPERIMENTAL PROCEDURES

Cloning and Mutagenesis. All clones were sequenced by the New England BioLabs (NEB) Core facility. All enzymes were obtained from NEB and used as described by the manufacturer. The gene encoding the Tko CDC21-1 intein and five native extein residues flanking each side of the element was synthesized by Invitrogen (Carlsbad, CA). The DNA was digested with *Xho*I and *Spe*I, purified from an agarose gel using the Wizard SV Gel and polymerase chain reaction (PCR) Clean-up System (Promega, Madison, WI), and ligated into similarly digested pMP1.[19] The resultant pMIP plasmid has the Tko CDC21-1 intein (I) flanked by the *Escherichia coli* maltose binding protein (M) and the Δ*Sal* fragment of *Dirofilaria immitis* paramyosin (P). The *Saccharomyces cerevisiae* Sce VMA

dx.doi.org/10.1021/bi201840k | *Biochemistry* 2012, 51, 2496−2505

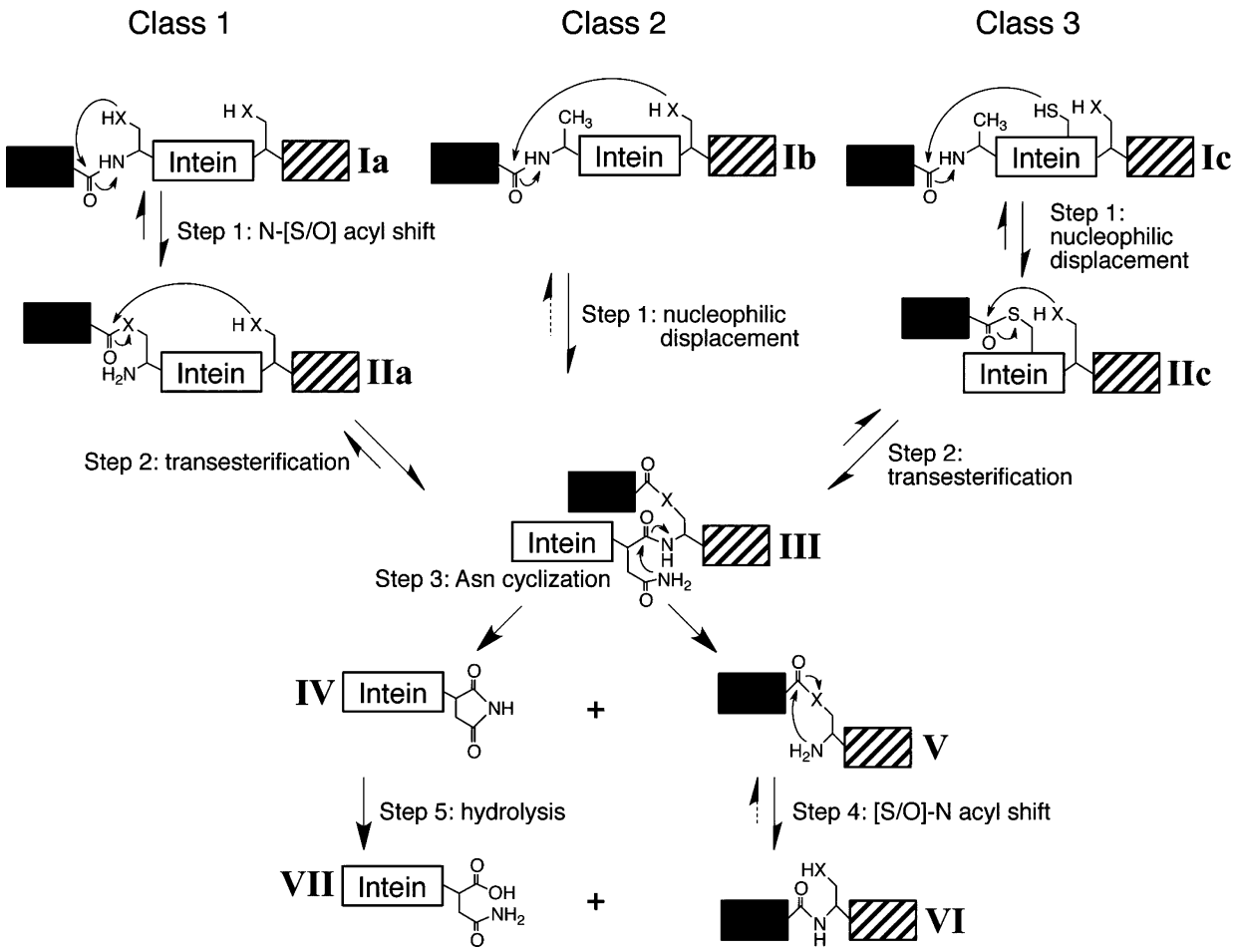


Figure 1. Splicing mechanism for the three classes of inteins. The majority of inteins follow the Class 1 intein-mediated protein splicing mechanism, which consists of four coordinated nucleophilic displacement reactions. Class 2 and Class 3 inteins form the Block G BI (**III**) by different pathways compared to those of Class 1 inteins because they generally start with amino acids incapable of forming the linear (thio)ester intermediate (**IIa**). In Class 2 inteins, Cys_{+1} is thought to directly attack an amide bond at the N-terminal splice junction to form the Block G BI (**III**). Class 3 inteins form the class-specific Block F BI (**IIc**) prior to the Block G BI (**III**). Once the Block G BI is formed, the remainder of the splicing reaction is the same in all classes of inteins. Residues within the intein assist these enzymatic reactions. However, in all classes, there is variability in the residues and positions within the intein that facilitate each reaction. Tetrahedral intermediates are not shown. The black rectangle represents the N-extein, the striped rectangle the C-extein, and X the sulfur or oxygen atom in the side chain of Ser, Thr, or Cys; solid arrows represent steps that have been experimentally verified.

intein gene[28] with codons for three native extein residues on both sides was amplified by PCR and cloned between M and P. The gene encoding the *Nostoc punctiforme* Npu DnaE intein[31] was synthesized by Invitrogen and cloned between M and P with three native extein residues on both sides. The Mja KlbA intein and *Mycobacterium xenopi* Mxe GyrA intein were previously cloned between M and P.[19,37] The Mja KlbA intein was flanked by a single native N-extein residue and three native C-extein residues.[19] The Mxe GyrA intein was flanked by 65 native N-extein residues and 64 native C-extein residues,[37] so its spliced MP product is ~125 aa larger than the spliced MP products from the other inteins. All mutants were constructed using the Phusion site-directed mutagenesis kit (NEB) with pMIP as the template and appropriate primers to introduce the desired mutation.

In Vivo and In Vitro Splicing. All MIP fusions were expressed in NEB Express I^q Competent *E. coli* (NEB) by induction at an OD_{600} of 0.5−0.7 with 0.4 mM isopropyl β-D-thiogalactoside (IPTG) in 10 mL of LB medium containing 100 μg/mL ampicillin for 2 h at 37 or 15 °C overnight. Harvested cells were disrupted by sonication in buffer A [20 mM Na_2HPO_4 (pH 7.0) and 500 mM NaCl]. Soluble lysates were heated for 1 h at 60 °C for in vitro splicing and then boiled for 5 min in SDS sample buffer with DTT (NEB). The samples were loaded onto 10 to 20% Tris-glycine polyacrylamide gels (Invitrogen) and then either stained with Simply Blue Safe Stain (Invitrogen) or transferred to nitrocellulose membranes for Western blotting with anti-maltose binding protein or anti-paramyosin antisera, as described previously,[19] to confirm the identity of M, IP, and the branched intermediate in the wild type and selected mutants, including T81A, T81H, C1A, D122A, K82A, T78A, and E80A, and all mutants with >10% N-terminal cleavage products (N140A and T+1A).

Amino Acid Sequence Alignment and Structural Modeling. Amino acid sequences from the Tko CDC21-1 intein and its four orthologs (*Haloquadratum walsbyi* Hwa MCM-2, *Pyrococcus abyssi* Pab CDC21-2, *Pyrococcus furiosus* Pfu CDC21, and *Pyrococcus horikoshii* Pho CDC21-2 inteins)[4] were aligned using the slow/accurate pairwise alignment with default options in ClustalW. N-terminal and C-terminal portions of the Tko CDC21-1 and Npu DnaE inteins were

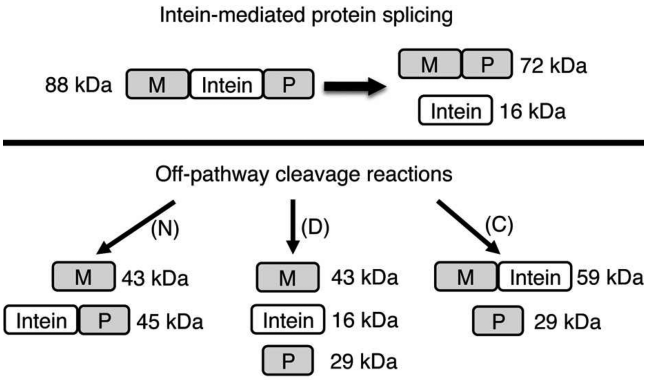


Figure 2. Splicing and off-pathway cleavage reactions for the Tko CDC21-1 intein in a model precursor. The MIP precursor consists of the maltose binding protein as the N-extein (M) and a fragment of paramyosin (P) as the C-extein along with five native extein residues flanking each side of the intein (I). Protein splicing yields ligated MP and free intein; off-pathway N-terminal splice junction cleavage (N) yields M + IP, and off-pathway C-terminal splice junction cleavage (C) yields MI + P. Cleavage of the thioester linkage in branched intermediates also yields M + IP. Some mutations result in double cleavage (D) at both splice junctions to yield M, I, and P. Molecular masses of the Tko CDC21-1 intein MIP precursor and potential products are given.

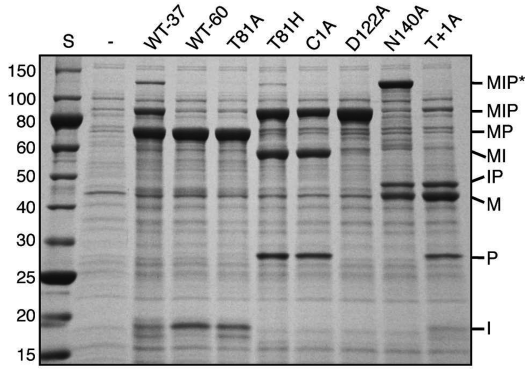


Figure 3. Splicing of wild-type (WT) and mutant Tko CDC21-1 inteins in a model precursor. MIP fusion proteins containing the indicated intein mutations were induced at 37 °C. Soluble protein from in vivo splicing of the WT intein (WT-37) or soluble protein from in vitro splicing of the wild-type (WT-60) and mutant intein samples after incubation at 60 °C for 1 h were electrophoresed via SDS–PAGE and stained with Simply Blue Safe Stain. Mutations are listed above each lane. The sizes of the molecular mass standards (S) are listed in kilodaltons (NEB, 10–250 kDa ladder). Abbreviations as in Figure 2 with the addition of MIP*, which is the branched intermediate, and lane –, which is the soluble extract from cultures containing the T81A mutant without induction by IPTG.

aligned separately using the same program and then appended. The sequence proximal to Block B in the Tko CDC21-1 intein and nine standard inteins with $His_{B:10}$ were manually aligned on the basis of known structures and sequences.[4,18,26−32] The Npu DnaE intein structure used to model the Tko CDC21-1 intein contained a C1A mutation and Gly-Gly as its N-extein (Protein Data Bank entry 2KEQ).[31] The alignment between the Npu DnaE intein and the Tko CDC21-1 intein with Gly-Leu native N-extein residues was submitted to the SWISS-MODEL web-based comparative modeling server in the alignment mode. The model was evaluated using ANOLEA and GROMOS as implemented in the SWISS-MODEL workspace, and ProSA-web.[38−40] The positions of Lys_{58} and Asp_{122} were modified to optimize the position of active residues for N-terminal scissile bond cleavage while avoiding steric clashes. The dihedral angles of Lys_{58} were adjusted using the standard rotomer tool in Swiss PDB.[41] The position of Lys_{58} was adjusted to 58.9° (Cα−Cβ), 180° (Cβ−Cγ), and 68.1° (Cγ−Cδ). The position of Asp_{122} was adjusted to 58.8° (Cα−Cβ) and 10° (Cβ−Cγ). All other side chain positions were unchanged.

RESULTS

Characterization of the Tko CDC21-1 Intein in a Model Precursor. A synthetic DNA encoding the Tko CDC21-1 intein with five native extein residues on each side (N-Extein, Ser-Ala-Ala-Gly-Leu; C-Extein, Thr-Ala-Ala-Ala-Val) was cloned in frame between the maltose binding protein (M) and the Δ*Sal* fragment of paramyosin (P) to form a MIP precursor. Spliced products (MP + I) were observed when MIP was expressed in *E. coli* at 37 °C for 2 h after induction with 0.4 mM IPTG (Figures 2 and 3 and Table S1 of the Supporting Information), which indicates that the Tko CDC21-1 intein is active even though it does not contain a His at position B:10. As some MIP precursor remained, expression at 15 °C overnight was tested because low-temperature expression often helps splicing in model precursors.[6,19−21] Some unspliced MIP precursor still remained when it was expressed at 15 °C with wild-type or mutated inteins (data not shown). Because *T. kodakaraensis* grows at 60–100 °C, splicing in vitro was tested at 60 °C. The remaining MIP precursor spliced to completion when soluble lysates from the 37 °C induction were heated at 60 °C for 1 h in vitro (Figure 3 and Table S1 of the Supporting Information). This heat treatment was performed in all subsequent Tko CDC21-1 intein experiments.

Ala and His substitutions were made to Thr_{81} at position B:10 in the Tko CDC21-1 intein (Table 1) to examine the importance of Thr_{81} for splicing (Figure 3 and Table S1 of the Supporting Information). T81A spliced completely, indicating that Thr_{81} does not play a significant catalytic role. On the other hand, neither spliced products (MP + I) nor off-pathway N-terminal cleavage products (M + IP) were observed when B:10 "reverted" back to the standard His residue. Instead, T81H yielded precursor (MIP) and C-terminal cleavage products (MI + P). The catalytic impediment caused by the His substitution indicates that the Tko CDC21-1 intein not only has evolved to splice without $His_{B:10}$ but also can no longer accommodate this large residue.

In Class 1 inteins, Cys_1 (A:1) is required for formation of the linear thioester intermediate (**IIa**), BI formation, and off-pathway N-terminal splice junction cleavage; the intein C-terminal Asn (G:7) is required for resolution of the BI (**III**) and off-pathway C-terminal splice junction cleavage, and the Thr_{+1} C-extein nucleophile (G:8) is required for BI formation and on-pathway N-terminal splice junction cleavage (Figure 1). The less conserved Asp in Block F (F:4) affects on-pathway N- and/or C-terminal cleavage.[26,30] The role of these catalytically important Class 1 intein residues was tested in the Tko CDC21-1 intein by introducing Ala substitutions at Cys_1, Asp_{122} (F:4), the intein C-terminal Asn_{140}, and Thr_{+1} (Figure 3 and Table S1 of the Supporting Information). C1A blocked N-terminal cleavage and yielded precursor (MIP) and C-terminal cleavage products (MI + P). D122A yielded MIP precursor. N140A blocked C-terminal cleavage, which resulted in accumulation of the BI (**III**, MIP*) and N-terminal cleavage

dx.doi.org/10.1021/bi201840k | *Biochemistry* 2012, 51, 2496–2505

products (M + IP). Ala substitution of Thr_{+1} blocked formation of the BI and yielded N-terminal cleavage products (M + IP), double-splice junction cleavage products (M, I, and P), and precursor (MIP). These results are consistent with the known properties of these essential residues and with splicing using the Class 1 intein pathway.

$Asp_{F:4}$ is known to be important for cleavage at both splice junctions and has been called a gatekeeper for C-terminal splice junction cleavage.[26,30,42] The role of Asp_{122} was examined in more detail by additional substitutions with Cys, His, Asn, or Glu (Table S1 and Figure S1 of the Supporting Information). D122C yielded mainly precursor (MIP) and trace amounts of C-terminal cleavage products (MI + P). His and Asn mutations blocked C-terminal cleavage, resulting in precursor (MIP) and a small amount of N-terminal cleavage products (M + IP). D122E was the only mutation that yielded a significant amount of spliced products (MP + I) along with smaller amounts of precursor (MIP), C-terminal cleavage products (MI + P), and double-cleavage products (M, I, and P). These results suggest that a carboxylic acid in the side chain of the residue at position F:4 is important for splicing in the Tko CDC21-1 intein.

Alanine Scanning of Block B, F, and G Residues To Identify Amino Acids That Activate the N-Terminal Splice Junction. Alanine scanning was performed on other residues in Blocks B, F, and G to determine if any of them were functionally substituting for the missing $His_{B:10}$. In Block B, Ser_{74} (B:3), Thr_{76} (B:5), Thr_{78} (B:7), Glu_{80} (B:9), Lys_{82} (B:11), and Thr_{85} (B:14) were mutated to Ala (Table 1 and Table S1 and Figure S2 of the Supporting Information). All of these Ala substitutions yielded >90% spliced product except T78A, indicating that none of these Block B residues except for Thr_{78} detectably assist splicing. The T78A mutation reduced the yield of spliced products by <50%. Thr_{78} is at position B:7, a position known to facilitate N-terminal cleavage to varying degrees in different inteins.

Alanine scanning was next performed in Block F at Thr_{124} (F:6), Glu_{126} (F:10), Ser_{128} (F:12), His_{129} (F:13), and Ser_{130} (F:14) and in Block G at His_{139} (G:6) (Table 1 and Table S1 and Figure S3 of the Supporting Information). Mutation of Glu_{126}, Ser_{128}, His_{129}, and Ser_{130} yielded >90% spliced products. The T124A mutation inhibited C-terminal cleavage and resulted in similar amounts of spliced products (MP + I) and the BI (**III**, MIP*), which is consistent with previous studies of the *Synechocystis* sp. Ssp DnaB intein.[32] Mutation of the intein penultimate His_{139} resulted in accumulation of the BI and a small amount of spliced products, precursor, and off-pathway N-terminal cleavage products (M + IP), which is consistent with its known role in assisting Asn cyclization and C-terminal cleavage. The remaining Tko CDC21-1 intein Block F residues are either small hydrophobic or aromatic residues (Table 1) and were not mutated in this study.

Next, residues outside the intein motifs were examined. Glu_{63} was positioned near the Tko CDC21-1 intein active site in the modeled structure described below, but mutations to Ala or Gln had no detectable effects on splicing (Table S1 of the Supporting Information). His_{117} precedes Block F and is the only remaining His in the Tko CDC21-1 intein not already tested. The H117A mutation did not affect splicing (Table S1 of the Supporting Information), demonstrating that no intein-encoded His is required for activation of the N-terminal splice junction.

Identification of a Conserved Tko CDC21-1 Ortholog-Specific Motif Containing a Residue That Is Important for Splicing. Alanine scanning in conserved motifs of the Tko CDC21-1 intein failed to identify a residue that replaced the missing contribution of $His_{B:10}$. Tko CDC21-1 intein orthologs were compared to look for ortholog-specific conserved residues or motifs that could functionally substitute for $His_{B:10}$ (Figure 4A and Figure S4 of the Supporting Information). Interestingly,

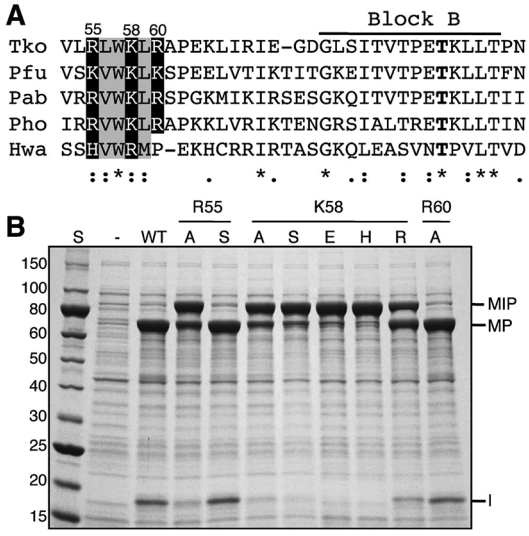


Figure 4. Importance of a conserved motif present in Tko CDC21-1 intein orthologs. (A) Multiple-amino acid sequence alignment of Block B and the preceding regions in Tko CDC21-1 intein orthologs. The basic and hydrophobic residues in the CDC21-1 ortholog-specific motif are shown with black and gray backgrounds, respectively. The B:10 Thr is shown in bold. The residue numbers in the Tko CDC21-1 intein are listed above the sequences. A consensus line below the sequence alignment uses the standard symbols generated and defined by ClustalW: *, identical residue in all sequences; :, highly conserved residue; ., weakly conserved residue. Abbreviations: Pfu, Pfu CDC21 intein; Pab, PAB CDC21-2 intein; Pho, Pho CDC21-2 intein; Hwa, Hwa MCM-2 intein. (B) Mutation of Tko CDC21-1 intein basic residues in the CDC21-1 ortholog-specific motif. MIP fusion proteins containing the indicated mutations were induced at 37 °C. Soluble proteins were incubated at 60 °C for 1 h in vitro and electrophoresed via SDS−PAGE followed by staining with Simply Blue Safe Stain. Mutations in residues Arg_{55}, Lys_{58}, and Arg_{60} are indicated above each lane. Abbreviations as in Figure 3.

multiple-sequence alignment revealed an ortholog-specific 6 aa motif preceding Block B consisting of basic residues at positions 1, 4, and 6, hydrophobic residues at positions 2 and 5, and Trp at position 3. Alanine substitutions were made at Arg_{55}, Lys_{58}, and Arg_{60} in the Tko CDC21-1 intein, which are at positions 1, 4, and 6, respectively, of this new motif (Figure 4 and Table S1 of the Supporting Information). R60A showed no significant difference in splicing compared to the wild-type intein. However, the level of splicing was drastically reduced in R55A and K58A. Since this inhibition might be due to changing a large hydrophilic residue to a small hydrophobic residue, a more conservative change to Ser was made at Arg_{55} and Lys_{58} (Figure 4B and Table S1 of the Supporting Information). R55S spliced to completion, which suggests that hydrophilicity is required at this site, not a basic residue. In contrast, K58S failed to splice and yielded products similar to those of K58A. Additionally, Met, Glu, Arg, and His mutations of Lys_{58} were introduced (Figure 4B and Table S1 of the Supporting Information). K58M, K58E, and K58H inhibited splicing and

dx.doi.org/10.1021/bi201840k | *Biochemistry* 2012, 51, 2496−2505

yielded mainly precursor. K58R was the only functional mutation at this position, resulting in splicing of >50% of the MIP precursor.

The New Tko CDC21-1 Intein Motif Is Not Universal. Sequence comparison indicated that the entire ortholog-specific motif was not conserved in nine inteins with known structures and $His_{B:10}$, but many have a basic residue preceded by an aromatic residue at a position similar to that of Lys_{58} in the Tko CDC21-1 intein (Figure S5A of the Supporting Information).[4,18,26,28−32] Ala substitutions were made to this basic residue in the Mja KlbA (Arg_{72}), Mxe GyrA (His_{52}), Sce VMA (Arg_{42}), and Npu DnaE (His_{48}) inteins. Splicing was examined at 37 or 15 °C in the Mja KlbA, Mxe GyrA, and Npu DnaE inteins and at 30 or 15 °C in the Sce VMA intein. All four inteins spliced to varying degrees at 37 or 30 °C (data not shown) and spliced to completion at 15 °C (Figure S5B of the Supporting Information). The Ala mutation at this basic residue had no observable effect on splicing in any of these four inteins at 15 °C (Figure S5B of the Supporting Information).

Comparative Modeling of the Tko CDC21-1 Intein. To provide insight into how Lys_{58} might facilitate splicing, a structural model of the Tko CDC21-1 intein was generated using SWISS-MODEL by threading it over the Npu DnaE intein NMR structure (Figure S6 of the Supporting Information).[31] The Npu DnaE intein has a relatively high level of similarity (29%) to the Tko CDC21-1 intein compared to other inteins with known structures. The overall protein model quality score (z score) using ProSA-web was satisfactory (z score = −4.26), but the model exhibited a few regions with positive ANOLEA or GROMOS energies, indicating minor defects in the local packing or electrostatic environment of some side chains. This is difficult to avoid when using a template with a low degree of sequence identity. Additional rotations of the side chains of Lys_{58} and Asp_{122} were performed to better approximate a catalytically functional active site. This model was used to constrain possible mechanisms by exploring whether residues could be positioned to support catalysis.

Standard active site residues and residues shown by mutation to be involved in activating the N-terminal splice junction scissile bond between Leu_{-1} and Cys_1 were examined, including Cys_1, Asp_{122} at F:4, Thr_{78} at B:7, Thr_{81} at B:10, and Lys_{58} (Figure 5A). Cys_1 was in position to perform the N−S acyl migration with the distance between its sulfhydryl and the Leu_{-1} carbonyl carbon being 3.1 Å. Asp_{122} was positioned to play a role in activating Cys_1 with a side chain oxygen (OD) 2.9 Å from the Cys_1 sulfur atom. The hydroxyl of Thr_{78} (B:7) was 3.3 Å from the amide nitrogen of Cys_1 and 2.9 Å from the Cys_1 carbonyl oxygen, positioned as in other inteins to stabilize the active site architecture. The Npu and Ssp DnaE intein structures show the B:10 His within hydrogen bonding distance of the scissile bond amide nitrogen.[18,31] However, the hydroxyl of Thr_{81} (B:10) in the Tko CDC21-1 intein model is 4.7 Å from the amide nitrogen of Cys_1 where it is clearly unable to interact with the splice junction amide nitrogen except possibly through a water molecule. These measurements verified that the positions of these active site residues in the modeled structure are consistent with their proposed catalytic functions and with the mutation data that show that Thr_{81} (B:10) does not detectably contribute to catalysis.

The modeled structure was next used to examine the position of Lys_{58}. The model predicted that the distance between the side chain NZ atom of Lys_{58} and the amide nitrogen or carboxyl oxygen of the N-terminal scissile bond was

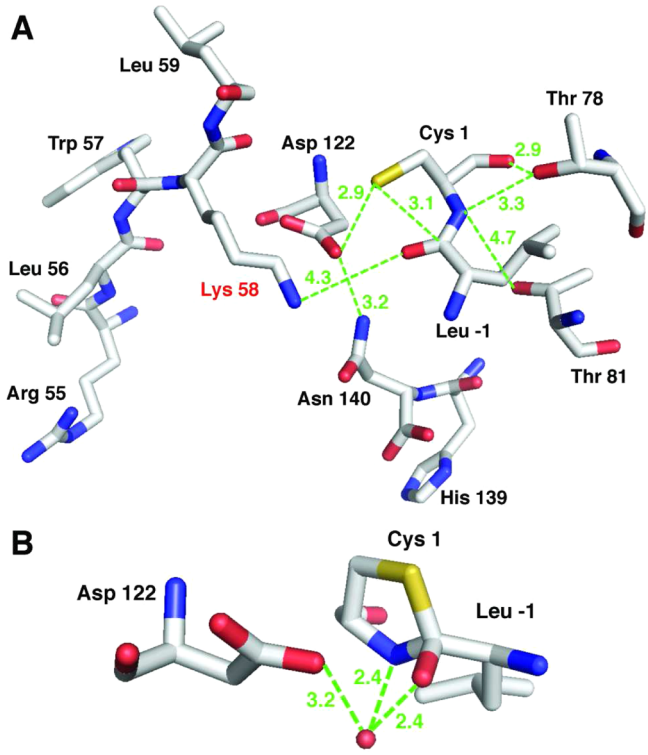


Figure 5. Active site region in the modeled structure of the Tko CDC21-1 intein. (A) The arrangement of residues near the splice junctions is depicted after the initial rotomer for Lys_{58} and Asp_{122} had been changed to bring the side chain of Lys_{58} into the active site and to remove the steric clash this caused with Asp_{122}. (B) Model of the proposed tetrahedral intermediate. The red sphere shows the position of a proposed water molecule. The backbone and side chain atoms of each residue are shown as sticks, with carbon atoms colored gray, nitrogen atoms blue, oxygen atoms red, and sulfur atoms yellow. Protons are not shown. Distances between atoms of neighboring side chains are represented by green dashed lines and labeled with the distance in angstroms.

6.6 or 4.3 Å, respectively. No rotomers of the side chain of Lys_{58} brought this residue within hydrogen bonding distance of the amide nitrogen at the N-terminal scissile bond.

DISCUSSION

Splicing of the Tko CDC21-1 Intein. $His_{B:10}$ is the most highly conserved intein residue,[4] and previous reports demonstrated that it is essential for splicing and off-pathway N-terminal splice junction cleavage in inteins from all three mechanistic classes.[6,7,19−21,30,33−35] Therefore, it was not surprising that the Arsp-FB24 Arth_1007 intein was unable to splice in the absence of a His at position B:10.[36] Remarkably, the Tko CDC21-1 intein spliced to completion in a model precursor even though $His_{B:10}$ is naturally replaced by a chemically dissimilar residue, Thr_{81}. Thr is conserved at position B:10 in all five Tko CDC21-1 intein orthologs currently in InBase.[4] However, mutating Thr_{81} to Ala did not detectably impair splicing, indicating that Thr_{81} is clearly not replacing the lost catalytic function normally provided by $His_{B:10}$. In all other respects, the Tko CDC21-1 intein followed the standard Class 1 protein splicing pathway as evidenced by observing the expected products after mutating the three essential Class 1 intein nucleophiles (Cys_1, Asn_{140}, and Thr_{+1}) and three known facilitating residues (Asp_{122} at F:4, Thr_{78} at

dx.doi.org/10.1021/bi201840k | Biochemistry 2012, 51, 2496−2505

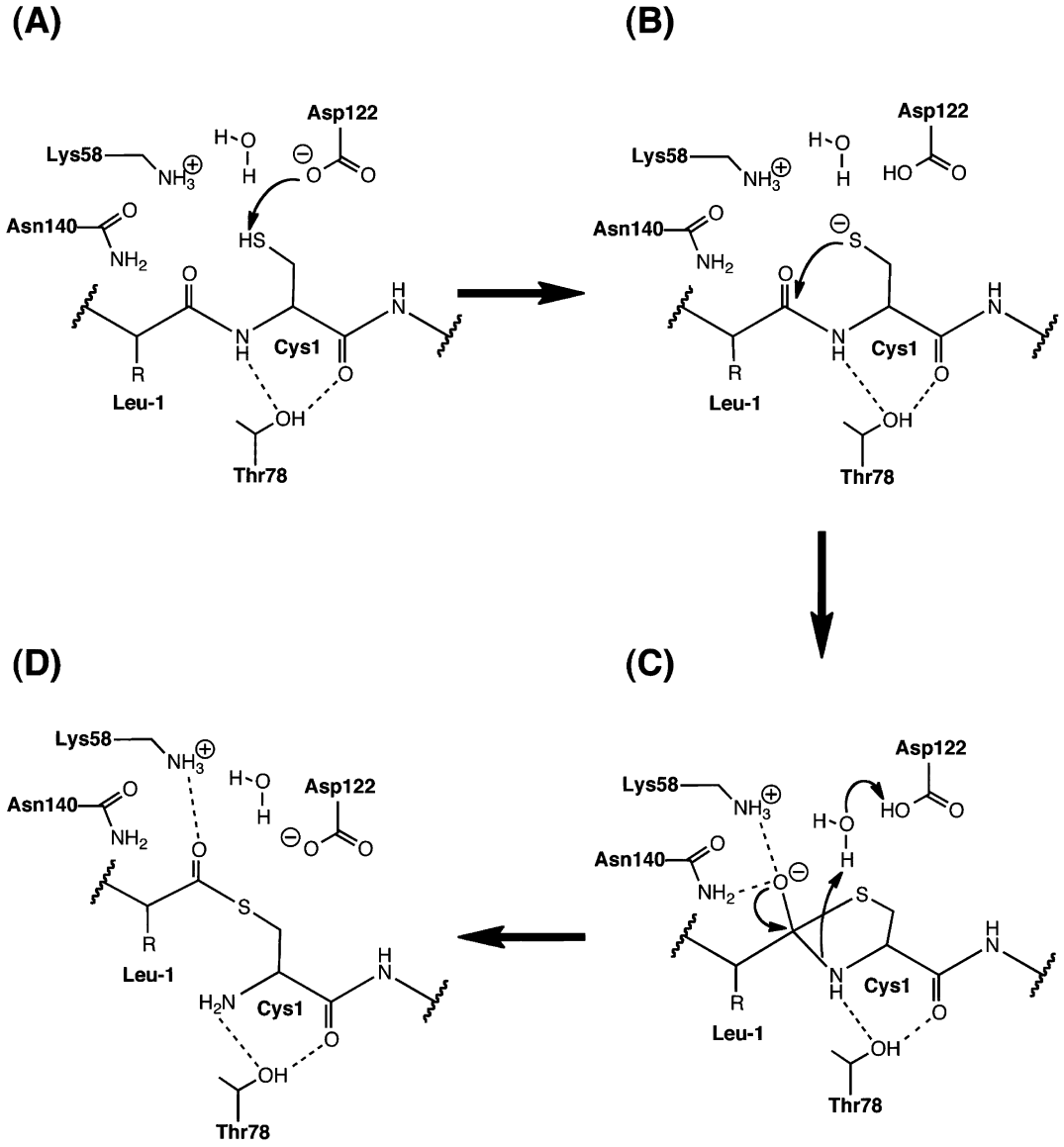


Figure 6. Proposed mechanism for the N−S acyl migration in the Tko CDC21-1 intein. (A) The N−S acyl migration is initiated by the deprotonation of the Cys_1 sulfur atom by Asp_{122}, making Cys_1 more nucleophilic. (B) Nucleophilic attack of the Cys_1 sulfur atom on the Leu_{-1} carbonyl carbon atom. Leu_{-1} is the N-extein residue at the N-terminal splice junction. (C) Tetrahedral intermediate of the initial acyl migration step. Lys_{58} of Tko CDC21-1 intein stabilizes the oxyanion on the tetrahedral intermediate along with Asn_{140}, while Asp_{122} (F:4) acts as a general acid to protonate the amide nitrogen leaving group through an intermediate water molecule. This protonation leads to decay of the tetrahedral intermediate, scissile bond cleavage, and formation of the thioester at the N-terminal splice junction. (D) Resultant thioester intermediate. During the process, Thr_{78} (B:7) forms a hydrogen bond with the carbonyl oxygen, the amide nitrogen of the scissile bond, or both, assisting the N−S acyl migration by positioning the scissile bond for catalysis.

B:7, and the penultimate His_{139} at G:7). Ala scanning of 10 more Block B, F, and G residues and the remaining histidine (His_{117}) in the Tko CDC21-1 intein failed to identify a candidate that might substitute for the absent $His_{B:10}$. However, alignment of the amino acid sequences of the five Tko CDC21-1 intein orthologs revealed a conserved motif outside the normal intein motifs. Ala substitution of the basic residues in this ortholog-specific motif demonstrated that a basic residue at position 4 (Lys_{58} in the Tko CDC21-1 intein) was critical for splicing and N-terminal cleavage and was thus a candidate for providing the functionalities lost by the absence of $His_{B:10}$.

Lys_{58} Does Not Perform the Proposed Acid−Base Catalysis Role of $His_{B:10}$. The exact role of $His_{B:10}$ in activating the N-terminal splice junction is unknown, but several hypotheses exist. One is that the initial N−[S/O] acyl migration is facilitated by ground state destabilization caused by a distorted N-terminal scissile bond,[27,29,33] which was shown in one case to require $His_{B:10}$.[33] However, not all intein structures show a distorted N-terminal scissile bond.[18,28,30−32] A second hypothesis states that $His_{B:10}$ is involved in general acid−base catalysis to promote the initial N−[S/O] acyl migration by protonating the amide nitrogen leaving group of the tetrahedral intermediate. Structures of the Sce VMA, Mxe GyrA, Ssp DnaE, Mja KlbA, Mtu RecA, and Ssp DnaB inteins show the ND1 atom of $His_{B:10}$ within hydrogen bonding distance of the amide nitrogen of the N-terminal scissile peptide bond.[18,26,27,29,30,32] Wang and co-workers proposed that $His_{B:10}$ plays a dual role in this first acyl shift, working as a base to activate the intein N-

dx.doi.org/10.1021/bi201840k | *Biochemistry* 2012, 51, 2496−2505

terminal nucleophile (Cys_1) and later serving as an acid to protonate the amide leaving group.[43]

These proposed mechanistic roles for $His_{B:10}$ can be ascribed to any residue with a similar side chain functionality and position within the intein structure. However, on the basis of the modeled structure of the Tko CDC21-1 intein (Figure 5A), no residue is positioned to protonate this leaving group. Although the side chain hydroxyl of Thr_{78} at position B:7 is within hydrogen bonding distance of the amide nitrogen of Cys_1, it most likely acts by positioning the active site, not in direct acid−base catalysis. The N-terminal scissile bond amide nitrogen faces away from Lys_{58}, and no rotomer of Lys_{58} can bring it close enough or in proper orientation to the amide nitrogen to facilitate protonation of the leaving group. Additionally, the biochemical data that show Lys_{58} can be successfully replaced with Arg strongly argue against an acid−base catalyst role for this residue. Instead, we propose that Lys_{58} stabilizes the tetrahedral intermediate in the initial N−S acyl shift at the N-terminal splice junction.

Proposed Mechanism for the N−S Acyl Migration in the Tko CDC21-1 Intein in the Absence of a His at Position B:10. Given the mutagenesis and structural modeling data, it is likely that the Tko CDC21-1 intein catalyzes the initial N−S acyl migration using a different atomic mechanism compared to that of the numerous inteins in which $His_{B:10}$ protonates the amide nitrogen leaving group. We propose the following mechanism for this reaction in the Tko CDC21-1 intein. Initially, Asp_{122} (F:4) facilitates deprotonation of the Cys_1 thiol to increase its nucleophilicity (Figure 6A); its side chain oxygen atom is 2.9 Å from the Cys_1 sulfur atom in the Tko CDC21-1 intein model structure (Figure 5A). Thr_{78} at position B:7 could assist this reaction by hydrogen bonding to both the scissile bond amide nitrogen (2.9 Å) and carbonyl oxygen (3.3 Å) to position Cys_1 in the active site.

The activated Cys_1 sulfur atom then attacks the carbonyl carbon of Leu_{-1}, the N-extein splice junction residue preceding the intein (Figure 6B), which results in formation of a thiazolidine−tetrahedral intermediate (Figure 6C). The oxyanion of the tetrahedral intermediate is transiently stabilized by a hydrogen bond with the amide side chain of the intein C-terminal Asn_{140} and by a charge−charge interaction (either directly or via a water molecule) with the ε-amine (NZ atom) of Lys_{58} (4.3 Å) (Figure 6C). The role of Lys_{58} is more critical to the N−S acyl migration than the role of Asn_{140} because Ala substitution of Asn_{140} still results in complete cleavage of the N-terminal splice junction, while Ala substitution of Lys_{58} results in mainly unreacted precursor (Figures 3 and 4). This proposed role for Lys_{58} is further supported by the observation that only replacement with another residue that can provide the same type of charge−charge interaction (K58R) yields significant spliced product and splicing is inhibited by the sterically similar Met substitution. A similar role of stabilizing the oxyanion of the tetrahedral intermediate has been proposed for $His_{B:10}$ in the Sce VMA intein.[28]

We propose that the transfer of a proton to the amide nitrogen of Cys_1 in the tetrahedral intermediate is accomplished by a catalytic water molecule located adjacent to Asp_{122} (3.2 Å) and the Cys_1 amide nitrogen (2.4 Å) (Figure 5B). This water molecule would play a role similar to that of $His_{B:10}$ in other inteins. Thus, the tetrahedral intermediate decays when a proton is passed directly from Asp_{122} to the Cys_1 amide nitrogen via the water intermediate, initiating cleavage of the scissile bond (Figure 6C). These proposed roles for Asp_{122} are supported by the mutation data that show conversion of Asp_{122} to Glu, a substitution that would allow proton transfer, permits splicing, but conversion of Asp_{122} to Asn, an isosteric side chain that is unlikely to perform the required acid−base chemistry, blocks splicing.

This scheme predicts that the Tko CDC21-1 intein has evolved a different atomic mechanism for catalysis because of the absence of a His at position B:10. The reactions of mutant inteins and the modeled intein structure both suggest that this intein has not substituted a chemically equivalent residue for $His_{B:10}$ that would increase the rate of decay of the thiazolidine intermediate by leaving group activation, thus shifting the microscopic equilibrium in favor of the linear thioester. Instead, the Tko CDC21-1 intein mitigates the loss of reactivity engendered by the lack of $His_{B:10}$ by stabilizing the thiazolidine intermediate using Lys_{58}. This allows the transfer of a proton from water, a far worse proton donor than His, to become feasible on a biological time scale.

In the intein world, the Tko CDC21-1 intein is not the only example in which a catalytically functional His is replaced with a residue that performs a chemically distinct function. Most active inteins naturally have a penultimate His, which normally facilitates the Asn cyclization step by organizing the active site and shuttling protons.[2,4,7−9,26−29,32] On the basis of the crystal structure of the Ssp DnaE intein that naturally has a penultimate Ala, Rao and co-workers predict that Arg_{73} at position B:11 stabilizes the oxyanion of the tetrahedral intermediate formed during the Asn cyclization step.[18] This predicted interaction is analogous to our proposal that Lys_{58} stabilizes the oxyanion of the N−S acyl migration tetrahedral intermediate in the Tko CDC21-1 intein. On the other hand, the nearly absolute conservation of $His_{B:10}$ is surprising, suggesting that it performs more than one important function. No alternatives to $His_{B:10}$ are found in nature except for the Tko CDC21-1 intein orthologs, which probably represent a single event that spread by lateral transmission.

Perspective. Throughout this discussion, the absence of a single, common strategy for performing each reaction step has exemplified the flexibility that inteins exhibit in performing the same set of nucleophilic displacement reactions. For example, examination of the proposed roles for $His_{B:10}$ in the initial acyl migration reveals three different strategies: ground state destabilization by scissile bond strain, protonation of the tetrahedral intermediate leaving group, and stabilization of the oxyanion of the tetrahedral intermediate. Self-catalytic systems, such as inteins, have an advantage over classical enzymes when exploring different strategies for performing the same overall enzymatic process. Because the enzyme−substrate complex is covalently tethered, it does not need rapid turnover to be biologically relevant. Therefore, despite having to coordinate multiple reaction steps to achieve successful maturation of the extein, inteins can more easily tolerate disruptive mutations that diminish the catalytic efficiency of one or more of these steps by utilizing second-site mutations to repair the catalytic defect. This has resulted in numerous variations in the intein-mediated protein splicing mechanism, such as the Class 3 inteins in which the absent intein N-terminal nucleophile is replaced with an internal Cys at position F:4 that cleaves the N-terminal splice junction. Analysis of the mechanistic details in various inteins can provide insight into how an enzyme can rescue activity after mutation and how it can catalyze production of the same molecule using different mechanisms within the same active site context and structural fold. As is often the case, we find that

dx.doi.org/10.1021/bi201840k | *Biochemistry* 2012, 51, 2496−2505

biology has discovered multiple unique solutions to solve the same chemical problems.

■ ASSOCIATED CONTENT

Supporting Information

Table S1 lists the products observed in all Tko CDC21-1 intein mutants from this study. The results of mutating residues in Tko CDC21-1 intein Blocks B, F, and G are presented in Figures S1−S3 as SDS−PAGE gels of soluble lysates from mutated MIP precursors. An amino acid sequence alignment and motif positions of Tko CDC21-1 intein orthologs is presented in Figure S4. The amino acid sequence alignment of the region equivalent to that of Lys_{58} of the Tko CDC21-1 intein is presented for nine inteins in Figure S5 along with the results of mutating this position. The sequence alignment between the Tko CDC21-1 intein and the Npu DnaE intein used for modeling is presented in Figure S6. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*New England BioLabs, 240 County Rd., Ipswich, MA 01938. Phone: (978) 380-7326. Fax: (978) 921-1350. E-mail: perler@neb.com.

Present Addresses

[†]Dana-Farber Cancer Institute, Boston, MA 02115, and The Broad Institute, Cambridge, MA 02142.

[‡]Emmanuel College, Boston, MA 02115.

Funding

This work was supported by New England BioLabs.

Notes

The authors declare no competing financial interest beyond the fact that the authors' employer, New England Biolabs, funded the project.

■ ACKNOWLEDGMENTS

We thank Don Comb (NEB) for support and encouragement. We thank Andy Gardner (NEB), Gregory Lohman (NEB), and Paul Balbo (NEB) for reading the manuscript and fruitful discussions.

■ ABBREVIATIONS

Arsp, *Arthrobacter* species; Tko, *T. kodakaraensis* KOD1; Mtu, *Mycobacterium tuberculosis*; Ssp, *Synechocystis* sp.; Mja, *M. jannaschii*; Mxe, *My. xenopi*; Sce, *S. cerevisiae*; Npu, *N. punctiforme*; Pfu, *P. furiosus*; Pab, *P. abyssi*; Pho, *P. horikoshii*; Hwa, *H. walsbyi*; MP-Be, mycobacteriophage Bethlehem; aa, amino acid; N-extein, amino-terminal extein fragment; C-extein, carboxy-terminal extein fragment; BI, branched intermediate; M, maltose binding protein; P, ΔSal fragment of *D. immitis* paramyosin; I, intein; MIP, model intein precursor consisting of the intein flanked by the maltose binding protein as the N-extein and the ΔSal fragment of *D. immitis* paramyosin as the C-extein; MP and I, spliced products of MIP; M + IP, products of N-terminal splice junction cleavage of MIP; MI + P, products of C-terminal splice junction cleavage of MIP; MIP*, branched intermediate of MIP; IPTG, isopropyl β-D-thiogalactoside; HINT, Hedgehog intein; WT, wild-type native sequence; DC, double cleavage at both splice junctions in the absence of extein ligation; NEB, New England BioLabs.

■ REFERENCES

(1) Liu, X. Q. (2000) Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu. Rev. Genet. 34*, 61−76.

(2) 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.

(3) Perler, F. B., Olsen, G. J., and Adam, E. (1997) Compilation and analysis of intein sequences. *Nucleic Acids Res. 25*, 1087−1093.

(4) Perler, F. B. (2002) InBase: The Intein Database. *Nucleic Acids Res. 30*, 383−384.

(5) Pietrokovski, S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci. 3*, 2340−2350.

(6) 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.

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

(8) Paulus, H. (2001) Inteins as enzymes. *Bioorg. Chem. 29*, 119−129.

(9) Xu, M. Q., and Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J. 15*, 5146−5153.

(10) Xu, M. Q., Comb, D. G., Paulus, H., Noren, C. J., Shao, Y., and Perler, F. B. (1994) Protein splicing: An analysis of the branched intermediate and its resolution by succinimide formation. *EMBO J. 13*, 5517−5522.

(11) Shao, Y., and Paulus, H. (1997) Protein splicing: Estimation of the rate of O-N and S-N acyl rearrangements, the last step of the splicing process. *J. Pept. Res. 50*, 193−198.

(12) Iwai, H., Zuger, S., Jin, J., and Tam, P. H. (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett. 580*, 1853−1858.

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

(14) Mills, K. V., Dorval, D. M., and Lewandowski, K. T. (2005) Kinetic analysis of the individual steps of protein splicing for the *Pyrococcus abyssi* PolII intein. *J. Biol. Chem. 280*, 2714−2720.

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

(16) Saleh, L., Southworth, M. W., Considine, N., O'Neill, C., Benner, J., Bollinger, J. M., and Perler, F. B. (2011) Branched Intermediate Formation Is the Slowest Step in the Protein Splicing Reaction of the Ala1 KlbA Intein from *Methanococcus jannaschii*. *Biochemistry 50*, 10576−10589.

(17) Nogami, S., Satow, Y., Ohya, Y., and Anraku, Y. (1997) Probing novel elements for protein splicing in the yeast Vma1 Protozyme: A study of replacement mutagenesis and intragenic suppression. *Genetics 147*, 73−85.

(18) Sun, P., Ye, S., Ferrandon, S., Evans, T. C., Xu, M. Q., and Rao, Z. (2005) Crystal structures of an intein from the split *dnaE* gene of *Synechocystis* sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *J. Mol. Biol. 353*, 1093−1105.

(19) 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.

(20) Brace, L. E., Southworth, M. W., Tori, K., Cushing, M. L., and Perler, F. (2010) The *Deinococcus radiodurans* Snf2 intein caught in the act: Detection of the Class 3 intein signature Block F branched intermediate. *Protein Sci. 19*, 1525−1533.

(21) Tori, K., and Perler, F. B. (2011) Expanding the definition of class 3 inteins and their proposed phage origin. *J. Bacteriol. 193*, 2035−2041.

(22) Mills, K. V., Manning, J. S., Garcia, A. M., and Wuerdeman, L. A. (2004) Protein splicing of a *Pyrococcus abyssi* intein with a C-terminal glutamine. *J. Biol. Chem. 279*, 20685−20691.
(23) Perler, F. B., Xu, M. Q., and Paulus, H. (1997) Protein splicing and autoproteolysis mechanisms. *Curr. Opin. Chem. Biol. 1*, 292−299.
(24) Teixeira, M. T., Fabre, E., and Dujon, B. (1999) Self-catalyzed cleavage of the yeast nucleoporin Nup145p precursor. *J. Biol. Chem. 274*, 32439−32444.
(25) Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) Crystal structure of a Hedgehog autoprocessing domain: Homology between Hedgehog and self-splicing proteins. *Cell 91*, 85−97.
(26) Van Roey, P., Pereira, B., Li, Z., Hiraga, K., Belfort, M., and 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*, 162−173.
(27) Poland, B. W., Xu, M. Q., and Quiocho, F. A. (2000) Structural insights into the protein splicing mechanism of PI-SceI. *J. Biol. Chem. 275*, 16408−16413.
(28) 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.
(29) 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.
(30) Johnson, M. A., Southworth, M. W., Herrmann, T., Brace, L., Perler, F. B., and Wuthrich, K. (2007) NMR structure of a KlbA intein precursor from *Methanococcus jannaschii*. *Protein Sci. 16*, 1316−1328.
(31) Oeemig, J. S., Aranko, A. S., Djupsjobacka, J., Heinamaki, K., and Iwai, H. (2009) Solution structure of DnaE intein from *Nostoc punctiforme*: Structural basis for the design of a new split intein suitable for site-specific chemical modification. *FEBS Lett. 583*, 1451−1456.
(32) 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.
(33) 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.
(34) Kawasaki, M., Nogami, S., Satow, Y., Ohya, Y., and Anraku, Y. (1997) Identification of three core regions essential for protein splicing of the yeast Vma1 Protozyme. A random mutagenesis study of the entire Vma1-derived endonuclease sequence. *J. Biol. Chem. 272*, 15668−15674.
(35) Perler, F. B. (2005) Protein splicing mechanisms and applications. *IUBMB Life 57*, 469−476.
(36) Tori, K., and Perler, F. B. (2011) The *Arthrobacter* Species FB24 Arth_1007 (DnaB) Intein Is a Pseudogene. *PLoS One 6*, e26361.
(37) Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W. R. Jr., and Perler, F. B. (1997) The *Mycobacterium xenopi* GyrA protein splicing element: Characterization of a minimal intein. *J. Bacteriol. 179*, 6378−6382.
(38) Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res. 37*, D387−D392.
(39) Melo, F., and Feytmans, E. (1998) Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol. 277*, 1141−1152.
(40) Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res. 35*, W407−W410.
(41) Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis 18*, 2714−2723.
(42) Pearl, E. J., Tyndall, J. D., Poulter, R. T., and Wilbanks, S. M. (2007) Sequence requirements for splicing by the Cne PRP8 intein. *FEBS Lett. 581*, 3000−3004.
(43) 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.