

# Structural, Dynamic, and Functional Characterization of a DnaX Mini-intein Derived from *Spirulina platensis* Provides Important Insights into Intein-Mediated Catalysis of Protein Splicing

Soumendu Boral, Snigdha Maiti, Aditya J. Basak, Woonghee Lee, and Soumya De\*

Cite This: *Biochemistry* 2020, 59, 4711–4724

Read Online

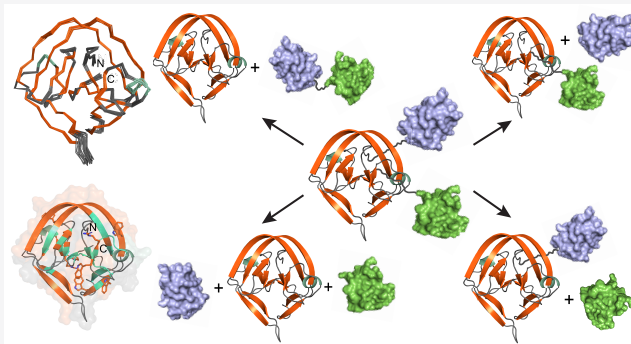
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Protein splicing is a self-catalyzed post-translational modification in which the intein enzyme excises itself from a precursor protein and ligates the flanking sequences to produce a mature protein. We report the solution structure of a 136-residue DnaX mini-intein enzyme derived from the cyanobacterium *Spirulina platensis*. This sequence adopts a well-defined globular structure and forms a horseshoe-shaped fold commonly found in the HINT (hedgehog intein) topology. Backbone dynamics and hydrogen exchange experiments revealed conserved motions on various time scales, which is proposed to be a characteristic of the intein fold. Interestingly, several dynamic motions were found in symmetrically equivalent positions within the protein structure, which might be a consequence of the symmetrical intein fold. In cell splicing activity showed that *Spl* DnaX mini-intein is a highly active enzyme. The precursor protein was not detected at any timepoint of the assay. Apart from the splicing reaction, catalytic cleavage at the N- and C-termini of the precursor protein was also observed. To determine the roles of the catalytic residues in splicing and cleavage reactions, all combinations of alanine mutations of these residues were generated and functionally characterized. This in-depth analysis revealed cooperativity between these catalytic residues, which suppresses the N- and C-terminal cleavage reactions and enhances the yield of the spliced product. Overall, this study provides a thorough structural, dynamic, and functional characterization of a new intein sequence and adds to the collection of these unique enzymes that have found tremendous applications in biochemistry and biotechnology.



Protein splicing is a unique post-translational rearrangement that was discovered in the early 1990s.<sup>1,2</sup> This process involves self-excision of an internal protein sequence, called an intein (intervening protein), from a precursor protein, resulting in the ligation of the two flanking N- and C-exteins (external proteins) into a single polypeptide chain via formation of a native peptide bond between them.<sup>3</sup> Since their discovery, intein enzymes have found numerous applications in chemical biology, protein engineering, biomedicine, and biotechnology.<sup>4,5</sup> A few examples of their applications are expressed protein ligation (EPL),<sup>6</sup> conditional protein splicing (CPS),<sup>7</sup> self-cleavage of the affinity tag for protein purification, site-specific chemical modification,<sup>8</sup> segmental isotope labeling,<sup>9</sup> and protein cyclization.<sup>10</sup> To date, >600 intein genes have been found and are deposited into the New England Biolabs Intein Database, InBase,<sup>11</sup> but only a few have been thoroughly characterized. Because inteins have a wide range of applications in protein chemistry, it is important to characterize more intein sequences that may find novel applications.

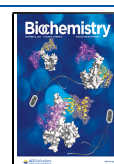
Protein splicing is an intramolecular reaction. Inteins are single-turnover enzymes and require neither cofactors nor energy sources. On the basis of their sequences and splicing

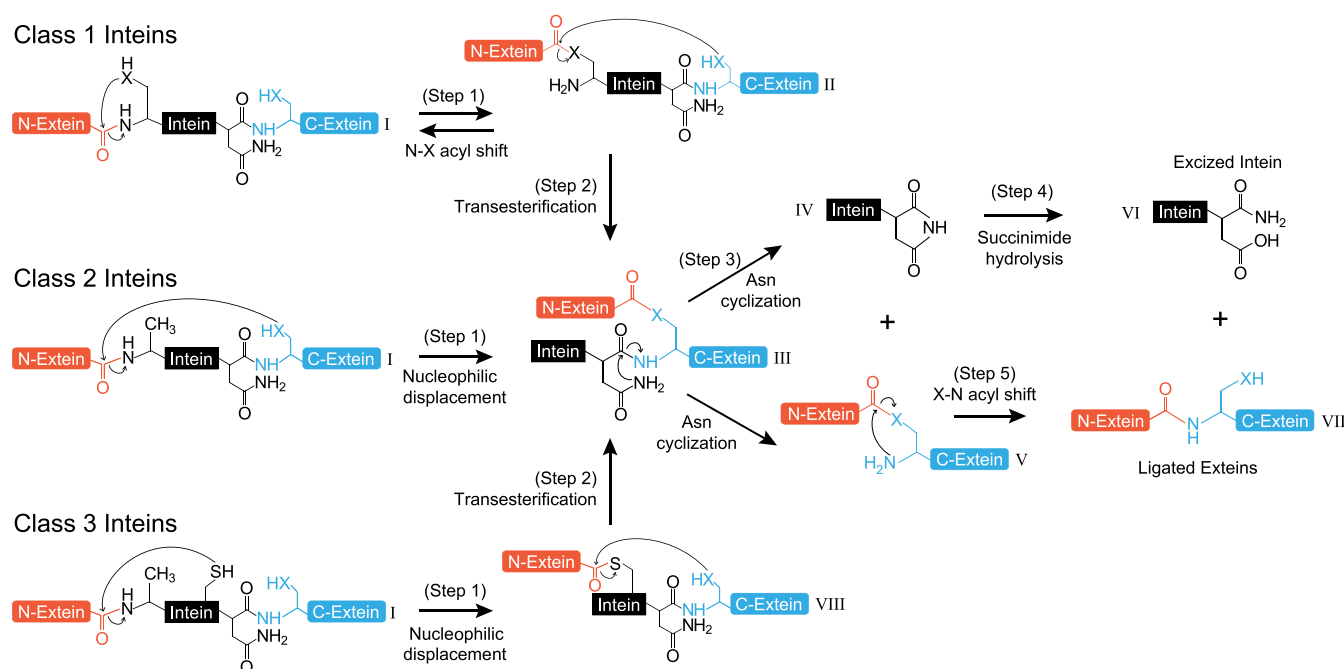
mechanisms, inteins are grouped into three classes.<sup>12</sup> For class 1 inteins, splicing is initiated by the N-terminal Cys or Ser side chain of the intein, which attacks the previous peptide bond, resulting in a thioester/ester linkage at the N-terminal splice junction (Figure 1). In the second step, the N-extein is then transferred to the side chain of the first C-extein residue (Cys, Ser, or Thr) by trans-(thio)esterification, resulting in a branched (thio)ester intermediate. In the third step, the side chain of the C-terminal asparagine attacks the following peptide bond, resulting in peptide bond cleavage, and separates the ester-linked exteins from the intein. Finally, the ester linkage rearranges to an amide linkage and the succinimide ring at the intein's C-terminal end hydrolyzes to form Asn.<sup>13</sup> This splicing reaction is catalyzed by a set of conserved active site residues of the intein enzyme.

Received: October 11, 2020

Revised: November 29, 2020

Published: December 8, 2020





**Figure 1.** Catalytic mechanisms of splicing reactions by three different classes of inteins. Class 1 protein splicing pathway consists of four nucleophilic rearrangement reactions, resulting in the spliced product (VII). The class 2 and class 3 inteins lack the N-terminal nucleophile. The first C-extein residue of class 2 inteins directly attacks the peptide bond at the N-terminal splice junction, resulting in the formation of the standard branched intermediate (III). In class 3 inteins, the side chain nucleophile of Cys in the WCT triplet attacks the N-terminal peptide bond to form a branched intermediate (VIII). The N-extein is then transferred to the side chain of the first C-extein residue, resulting in the formation of the same branched intermediate (III) as in class 1 and 2 inteins. Once the standard branched intermediate is formed, the remaining steps are the same in all three classes of inteins. The precursor protein, linear ester intermediate, cyclized intein, and ester-linked exteins are shown as I, II, IV, and V, respectively. X represents the sulfur or oxygen atom in the side chain of Cys, Ser, or Thr. Tetrahedral intermediates are not shown here.

However, both class 2 and class 3 inteins lack the N-terminal Cys or Ser nucleophile; thus, they are unable to form the linear thioester intermediate (Figure 1). In the case of class 2 inteins,<sup>14</sup> the side chain of the first C-extein residue directly attacks the N-terminal scissile bond, thereby omitting both the acyl rearrangement and trans-esterification reaction and resulting in the formation of the branched intermediate as in class 1 inteins. Class 3 inteins contain a conserved Trp-Cys-Thr (WCT) triplet, whose Cys side chain attacks the N-terminal scissile amide bond, resulting in a branched intermediate.<sup>15</sup> In the second step, the N-extein is transferred from the side chain of this Cys to the side chain of the first C-extein residue to form another branched intermediate, similar to class 1 and 2 inteins. Once the standard branched intermediate is formed, both class 2 and class 3 inteins follow the same class 1 splicing pathway.<sup>4</sup>

In this study, we have carried out the structural and functional characterization of a new intein sequence using solution NMR spectroscopy and an in cell splicing assay, respectively. The 136-residue intein was derived from  $\gamma$  and  $\tau$  subunits of the DNA polymerase III (DnaX) gene of a cyanobacterium *Spirulina platensis*, abbreviated as *Spl* DnaX intein.<sup>11</sup> This intein is classified as a mini-intein due to the absence of a homing endonuclease domain. We have determined the structure of *Spl* DnaX mini-intein by solution NMR spectroscopy. Although the level of sequence identity with existing intein structures is low, *Spl* DnaX mini-intein is found to form a horseshoe-shaped fold commonly found in the HINT (hedgehog/intein) domain superfamily.<sup>16–19</sup> NMR-based hydrogen exchange experiments

reveal that the core of the intein enzyme is very stable. Backbone dynamics experiments show the presence of fast (picoseconds to nanoseconds) and slow (microseconds to milliseconds) time scale motions, which appear to be conserved in several intein enzymes.<sup>20–23</sup> Also, dynamic motions have been detected in symmetric positions within the intein structure, most likely a consequence of the symmetrical fold of the protein. The enzyme was found to be highly functional. The splicing reaction, within the bacterial cell, was monitored by gel electrophoresis of the whole cell lysate, and the splicing and cleavage products were detected by Western blotting analysis. On the basis of the determined structure, the catalytically important residues were identified and their functional role in the catalytic mechanism was determined by site-directed mutagenesis. Overall, in this study, we show that the *Spl* DnaX mini-intein has a well-folded and rigid three-dimensional structure, is highly active, and follows the class 1 splicing mechanism.

## MATERIALS AND METHODS

**Cloning, Protein Expression, and Purification.** A gene construct corresponding to the splicing domain (residues Glu129–Cys266) of an intein DnaX (UniprotKB K1XG21) from *S. platensis* was cloned into the pET28a(+) vector. Residue Glu129 is numbered as E(–1). The plasmid was transformed into the *Escherichia coli* BL21( $\lambda$ DE3) strain. For the <sup>15</sup>N-labeled protein sample, 1 L of M9 minimal medium was supplemented with 1 g of [<sup>15</sup>N]ammonium chloride and 5 g of [<sup>12</sup>C]glucose as the sole nitrogen and carbon sources, respectively. The culture

**Table 1. List of *Spl* DnaX Mini-intein Constructs**

name	description
intein <sup>WT</sup>	Residues before C1 and after N136 are spontaneously cleaved off due to its enzymatic action. Total of 136 residues.
intein <sup>C1A/C(+1)A</sup>	Both C1 and C(+1) are mutated to alanine. The residues GSHM prior to E(−1) are left after thrombin cleavage. Total of 141 residues.
UBQ-INT <sup>WT</sup> -HD	Fusion of N- and C-termini of <i>Spl</i> DnaX intein protein with ubiquitin (N-extein) and HD (C-extein) protein, respectively. The N-terminus has a His <sub>6</sub> tag, and the C-terminus has a 3X-Flag tag.
UBQ-INT <sup>C1A</sup> -HD	Containing a mutation of C1 to alanine in the UBQ-INT-HD construct
UBQ-INT <sup>N136A</sup> -HD	Containing a mutation of N136 to alanine in the UBQ-INT-HD construct
UBQ-INT <sup>C(+1)A</sup> -HD	Containing a mutation of C(+1) to alanine in the UBQ-INT-HD construct
UBQ-INT <sup>C1A/N136A</sup> -HD	Containing mutations of C1 and N136 to alanine in the UBQ-INT-HD construct
UBQ-INT <sup>C1A/C(+1)A</sup> -HD	Containing mutations of C1 and C(+1) to alanine in the UBQ-INT-HD construct
UBQ-INT <sup>N136A/C(+1)A</sup> -HD	Containing mutations of N136 and C(+1) to alanine in the UBQ-INT-HD construct
UBQ-INT <sup>C1A/N136A/C(+1)A</sup> -HD	Containing mutations of C1, N136, and C(+1) to alanine in the UBQ-INT-HD construct

was grown at 37 °C until cell density OD<sub>600</sub> reached ~0.7. After being induced with 0.5 mM IPTG for 5 h at 37 °C, cells were harvested and lysed by sonication in a lysis buffer [100 mM Tris-HCl, 200 mM NaCl, and 10 mM imidazole (pH 8.2)]. The wild-type intein cleaved off the N-terminal His<sub>6</sub> tag, thus hampering affinity purification. Hence, after cell lysis, the lysate was centrifuged at 16000g for 50 min to precipitate the cell debris. The supernatant was heated to 70 °C for 15 min and centrifuged to precipitate unfolded protein aggregates. The resulting clarified supernatant was filtered and passed through anion and cation exchange resins, respectively. Finally, size exclusion chromatography was performed to obtain highly pure intein protein.

For structural studies, a catalytically dead mutant intein was designed by mutating C1 and C(+1) to alanine. The resulting construct, intein<sup>C1A/C(+1)A</sup> (Table 1), was overexpressed in 1 L of M9 minimal medium supplemented with 1 g of [<sup>15</sup>N]-ammonium chloride and 3 g of [<sup>13</sup>C]glucose and purified using a Ni-NTA column. The Thrombin CleanCleave Kit (Sigma-Aldrich) was used to remove the affinity tag. The protein was exchanged into the final buffer [20 mM sodium phosphate and 50 mM NaCl (pH 6.5)]. The protein concentration was determined by ultraviolet absorption using the predicted molar absorptivity ( $\epsilon_{280}$ ). The protein concentrations were 0.6–0.7 mM with 7% lock D<sub>2</sub>O. For the long-term stability of the proteins, 0.8 mM PMSE, 2  $\mu$ L of protease inhibitor cocktail, and 0.04% sodium azide were also added to the final sample.

**Analytical Size Exclusion Chromatography (SEC).** Purified intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup> in a buffer containing 20 mM Tris (pH 7.5) and 50 mM NaCl were subjected to a prepacked Bio-Rad ENrich SEC 70 size exclusion column. Analytical SEC profiles were compared to probe any hydrodynamic shape difference between the two protein constructs.

**Analytical Ultracentrifugation (AUC).** The sedimentation velocity analytical ultracentrifugation (SV-AUC) experiment was performed on intein<sup>WT</sup> and the double mutant intein<sup>C1A/C(+1)A</sup> using a ProteomeLab XL-1 analytical ultracentrifuge (Beckman Coulter). Epon double-sector centerpieces were filled with 420  $\mu$ L of the sample buffer [20 mM Tris and 50 mM NaCl (pH 7.5)] and 400  $\mu$ L of intein protein having an  $A_{280}$  value of 0.4–0.6 in the two sectors. The samples were centrifuged at 40000 rpm and 20 °C. Frames were collected until sedimentation was complete. Absorbance scans were acquired in 3 min intervals. The sample buffer density, viscosity, and partial specific volume of the proteins were calculated by SEDNTERP;<sup>24</sup> data analysis was performed by SEDFIT using  $c(s)$  distribution analysis,<sup>25</sup> and figures were prepared using GUSSI.<sup>26</sup>

**NMR Experiments and Structure Calculations.** NMR experiments were performed in a 600 MHz Bruker Avance III spectrometer equipped with a triple-resonance cryogenic probe head. All studies were performed at 30 °C unless otherwise stated. For the backbone and side chain assignments, the <sup>15</sup>N, <sup>13</sup>C intein<sup>C1A/C(+1)A</sup> protein sample doubly labeled with <sup>15</sup>N and <sup>13</sup>C was used. Backbone resonances were manually assigned in NMRAM-SPARKY 1.470<sup>27</sup> using two-dimensional <sup>15</sup>N–<sup>1</sup>H HSQC and three-dimensional (3D) NMR experiments such as CBCA(CO)NH, CBCANH, HNCO, HN(CA)CO, and HNCA. Side chains were assigned using HBHA(CO)NH, H(CC)(CO)NH, (H)CC(CO)NH, HCCH-TOCSY, and <sup>15</sup>N-TOCSY experiments.<sup>28</sup> These spectra were processed and analyzed using NMRPipe<sup>29</sup> and Sparky,<sup>30</sup> respectively. The secondary structure propensity of this protein was calculated from the backbone chemical shifts (<sup>1</sup>H<sup>N</sup>, <sup>15</sup>N, <sup>13</sup>C $_{\alpha}$ , <sup>13</sup>C $_{\beta}$ , and <sup>13</sup>C $_{\gamma}$ ) using MICS.<sup>31</sup> For the structure calculation, a <sup>15</sup>N-edited NOESY spectrum with a mixing time of 110 ms and a sensitive enhanced <sup>13</sup>C-edited NOESY spectrum with a mixing time of 110 ms were recorded.

Chemical shift-based torsion angle restraints and nuclear Overhauser effect (NOE)-derived distance restraints were used to calculate the 3D structure of the intein. Initially, well-resolved NOE cross-peaks were assigned manually, and fold calculations were performed with PONDEROSA.<sup>32</sup> Xplor-NIH-based calculations<sup>33</sup> implemented in AUDANA<sup>34</sup> were used for the automated NOE assignment in the PONDEROSA-C/S suite.<sup>35</sup> Structure calculations were carried out by submitting jobs to the Ponderosa web server.<sup>36</sup> After several iterative calculations, the final 100 structures were determined with explicit water refinement. The final structures were validated using the wwPDB Validation Service (<https://validate-rcsb-2.wwpdb.org/>)<sup>37</sup> and PSVS 1.5<sup>38</sup> online servers. Secondary structural boundaries were determined using DSSP,<sup>39</sup> and structures were generated with PyMOL (PyMOL molecular graphics system, version 2.0.4, Schrodinger, LLC, New York, NY). Structural NMR statistics for the top 15 of 100 lowest energetically stable conformers are summarized in Table 2.

**Backbone Amide <sup>15</sup>N Relaxation.** Amide <sup>15</sup>N relaxation ( $R_1$ ,  $R_2$ , and steady-state heteronuclear NOE) experiments were performed at 30 °C for intein<sup>C1A/C(+1)A</sup> protein as described previously.<sup>40,41</sup> Briefly,  $R_1$  (50, 100, 200, 300, 400, 500, 750, 1000, and 1200 ms) and  $R_2$  (20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 ms) spectra were collected in random order to minimize any systematic error. The peak intensities of each residue were fit to the exponential decay function [ $I_t = I_0 \times \exp(-tR_i)$ ], where  $I_t$  is the peak intensity,  $t$  is the relaxation



**Table 2. Experimental Data and Statistics for the Ensemble of the 15 Best Solution NMR Structures of *Spl* DnaX Intein**

(a) NMR distance and dihedral restraints	
distance restraints	
total no. of NOEs	2108
short-range ( $ i - j  \leq 1$ )	984
medium-range ( $1 <  i - j  < 5$ )	220
long-range ( $ i - j  \geq 5$ )	904
hydrogen bond restraints	56
dihedral angle restraints	
total	238
$\varphi$	117
$\psi$	121
total no. of restricting constraints	2402
(b) average RMSD (Å) against the lowest-energy conformer for ordered residues <sup>a</sup>	
backbone atoms (N, C $^{\alpha}$ , C, O)	0.277
all heavy atoms	0.528
(c) RMSD from ideal geometry	
bond lengths (Å)	0.015
bond angles (deg)	1.4
(d) violations	
distance constraints (>0.5 Å)	0
dihedral angle constraints (>5°)	0
van der Waals constraints (>0.2 Å)	0
(e) Ramachandran plot statistics (%) from PROCHECK for selected residues <sup>b</sup>	
residues in most favored regions	90.3
residues in additionally allowed regions	9.5
residues in generously allowed regions	0.2
residues in disallowed regions	0
(f) Ramachandran plot statistics (%) from MolProbity for selected residues <sup>c</sup>	
residues in most favorable regions	97.1
residues in additionally allowed regions	2.9
residues in disallowed regions	0

<sup>a</sup>Ordered CYRANGE residues 1–137. <sup>b</sup>Selected residues 2–66 and 69–136. <sup>c</sup>With respect to the mean and SD for a set of 252 X-ray structures with <500 residues, of resolution  $\leq 1.80$  Å, R-factor of  $\leq 0.25$ , and R-free of  $\leq 0.28$ ; a positive value indicates a “better” score. Structure quality was evaluated using wwPDB validation pipeline (wwPDB VP, 2.12) and PSVS 1.5.

delay,  $I_0$  is the initial intensity, and  $R_i$  is either  $R_1$  or  $R_2$ . Errors in the calculated rate constants were determined by the Monte Carlo simulation. The ratio of the peak heights for each residue in the NOE versus a reference spectrum determined the heteronuclear  $\{^1\text{H}\}-^{15}\text{N}$  NOE. Uncertainties in the heteronuclear NOE were estimated by the propagation of error using spectral noise using the formula

$$\Delta\text{NOE} = \{[(1/I_{\text{REF}})\delta I_{\text{NOE}}]^2 + [(-I_{\text{NOE}}/I_{\text{REF}})^2\delta I_{\text{REF}}]^2\}^{1/2}$$

where  $\Delta\text{NOE}$  is the propagation of error in the heteronuclear NOE,  $I_{\text{NOE}}$  and  $I_{\text{REF}}$  are the peak heights of the NOE and control reference spectrum, respectively, and  $\delta I_{\text{NOE}}$  and  $\delta I_{\text{REF}}$  are the spectral noise of the NOE and control reference spectrum, respectively. The residue-specific order parameter ( $S^2$ ) was calculated using TENSOR2.<sup>42</sup>

**Hydrogen Exchange Experiments.** Amide proton–deuterium exchange rates were measured at 25 °C and pH 6.5. The protein was lyophilized and dissolved in 100% D<sub>2</sub>O, and a series of  $^{15}\text{N}-^1\text{H}$  HSQC spectra were collected over a period of time to monitor the decay of the amide signals as the

protium is exchanged by deuterium. The pseudo-first-order rate constants for exchange,  $k_{\text{ex}}$ , were calculated using in-house MATLAB codes by nonlinear least-squares fitting of the peak intensities,  $I_t$  (normalized by the number of transients) to the equation  $I_t = (I_0 - I_{00}) \times \exp(-k_{\text{ex}}t) + I_{00}$ , where  $t$  is the midpoint time of each spectrum,  $I_0$  is the initial peak intensity, and  $I_{00}$  accounts for the intensity coming from residual water.<sup>43,44</sup> The error in  $k_{\text{ex}}$  was determined by Monte Carlo simulation.

Amide proton–proton exchange rates were measured at pH 6.5 and 7.5 by the CLEANEX-PM method<sup>45</sup> at 30 °C. At each pH, a series of spectra with 10, 20, 30, 40, 50, 60, and 80 ms transfer periods and a reference spectrum using a recycle delay of 12.0 s were collected. The pseudo-first-order rate constants for chemical exchange,  $k_{\text{ex}}$ , were calculated by nonlinear least-squares fitting of the peak intensities versus transfer time using in-house MATLAB codes. A scaling factor of 0.7 was used to correct for the steady-state water magnetization.

The protection factors (PFs) for each amide proton were calculated as the ratio of the predicted intrinsic exchange rate constant ( $k_{\text{int}}$ ) for an unstructured polypeptide with the same amino acid sequence versus the experimentally calculated exchange rate constant ( $k_{\text{ex}}$ ). The  $k_{\text{int}}$  values were calculated using Sphere (<https://protocol.fccc.edu/research/labs/roder/sphere/sphere.html>).<sup>46</sup> Here an EX<sub>2</sub> mechanism is assumed, where the exchange rate constants ( $k_{\text{ex}}$ ) have a first-order dependence on sample pH and temperature.<sup>47</sup> Also, protein stability is assumed to be independent of pH and temperature.

**Splicing Activity Assay of Wild-Type and Mutant Inteins.** The 5′ and 3′ ends of *Spl* DnaX mini-intein gene were stitched by PCR with a ubiquitin gene (N-extein) and a homeodomain gene (C-extein) of HOX proteins, respectively (Figure S1). This was cloned into the pET28a(+) vector, resulting in an N-terminal His<sub>6</sub> tag. Additionally, a 3×-Flag tag was also added to the C-terminus of this construct. Several point mutations were introduced into the intein by inverse PCR-based site-directed mutagenesis. These plasmids containing His<sub>6</sub>-UBQ-Intein-HD-Flag<sub>3</sub> were transformed into *E. coli* BL21-(λDE3) cells and grown in LB medium supplemented with 70 μg/mL kanamycin at 37 or 16 °C, as indicated, until cell density OD<sub>600</sub> reached ~0.6–0.7. Protein expression was induced by 1 mM IPTG. Different time points were collected up to 1 h post-induction. Collected cells were harvested by centrifugation and immediately lysed with 1× SDS–PAGE loading dye containing 20% BME. The samples were boiled for ~10 min and centrifuged, and the supernatant from the lysate was subjected to 12% SDS–PAGE.

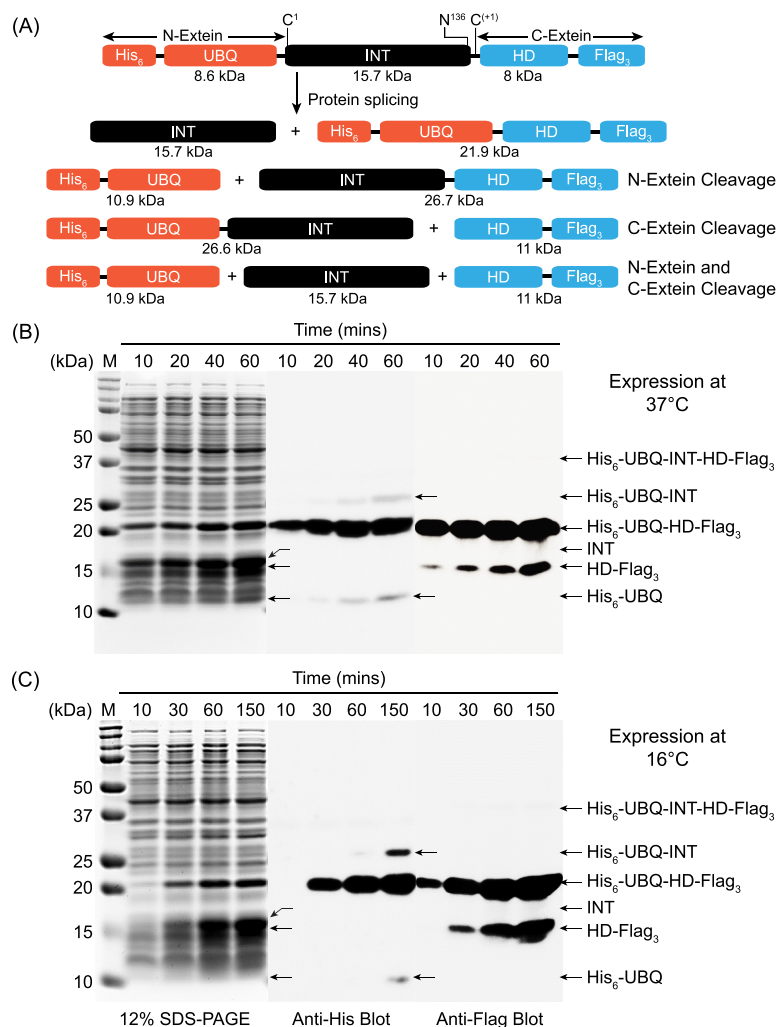
For Western blot analysis, the gels were blotted onto a PVDF (polyvinylidene difluoride) membrane with a pore size of 0.45 μm at 20 V for 20 min. The blots were incubated for 1 h in 5% nonfat dry milk powder solution in 1× TBST (Tris-buffered saline with Tween 20) at 25 °C to eliminate any nonspecific binding and then with a 1:5000 dilution of His<sub>6</sub>-tag and Flag-tag mouse monoclonal antibody, separately, in 1× TBST for 16 h at 4 °C. The blots were washed three times with TBST at 10 min intervals, incubated with a 1:10000 dilution of the goat anti-mouse antibody for 1 h at 25 °C, and then washed again three times with TBST in 10 min intervals. The blots were developed for 5 min with chemiluminescent substrate Luminol for horseradish peroxidase (HRP). A comparison with the protein marker allowed the identification of bands from each product. The band intensities were determined using ImageJ.<sup>48</sup> The yields of the splicing and the cleavage products were normalized



**Table 3. Relative Quantification of the Splicing and Cleavage Products of the Wild-Type and Mutant *Spl* DnaX Inteins<sup>a</sup>**

name	N-terminal cleavage	C-terminal cleavage	spliced product	precursor protein
UBQ-INT <sup>WT</sup> -HD	21.6%	19.8%	58.6%	ND <sup>b</sup>
UBQ-INT <sup>C1A</sup> -HD	ND <sup>b</sup>	90.5%	ND <sup>b</sup>	9.5%
UBQ-INT <sup>N136A</sup> -HD	7.5%	ND <sup>b</sup>	ND <sup>b</sup>	92.5%
UBQ-INT <sup>C(+1)A</sup> -HD	0.5%	67.8%	ND <sup>b</sup>	31.7%
UBQ-INT <sup>C1A/N136A</sup> -HD	ND <sup>b</sup>	ND <sup>b</sup>	ND <sup>b</sup>	100%
UBQ-INT <sup>C1A/C(+1)A</sup> -HD	ND <sup>b</sup>	70.2%	ND <sup>b</sup>	29.8%
UBQ-INT <sup>N136A/C(+1)A</sup> -HD	0.2%	ND <sup>b</sup>	ND <sup>b</sup>	99.8%
UBQ-INT <sup>C1A/N136A/C(+1)A</sup> -HD	ND <sup>b</sup>	ND <sup>b</sup>	ND <sup>b</sup>	100%

<sup>a</sup>Quantification of the bands is based on anti-His<sub>6</sub> Western blot images at induction for 60 min. <sup>b</sup>Not detected.



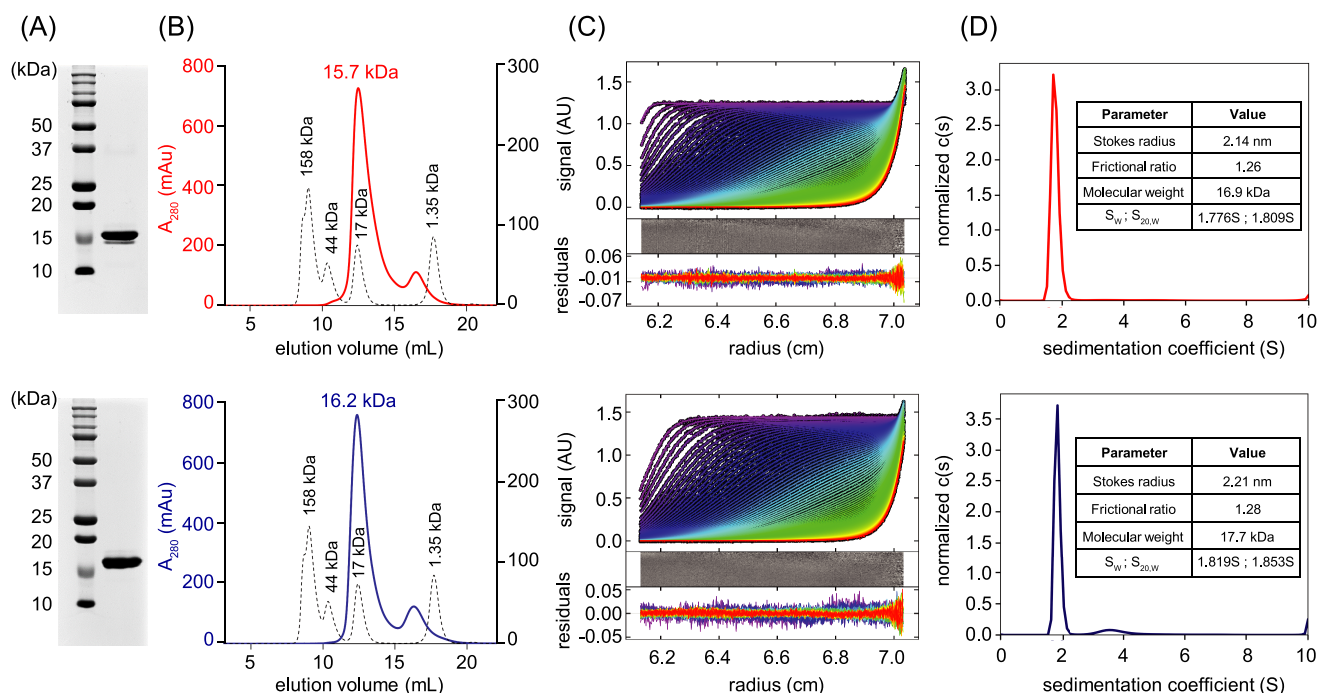
**Figure 2.** *Spl* DnaX mini-intein is catalytically active. (A) Schematic representation of the intein construct in which *Spl* DnaX mini-intein (INT) is fused with ubiquitin (UBQ) and homeodomain (HD) as the N-extein and C-extein, respectively. The N- and C-termini have His<sub>6</sub> and 3×-Flag tags, respectively. The positions of the three catalytically important residues, Cys1, Asn136, and Cys(+1), are shown. The spliced products and cleavage products due to side reactions are also indicated with their respective molecular weights. Protein was expressed at (B) 37 °C or (C) 16 °C. After induction, the *E. coli* cells were harvested at different times, as indicated, and analyzed by 12% SDS-PAGE. Each sample was run in triplicate and analyzed by Coomassie blue staining and anti-His<sub>6</sub> and anti-Flag Western blots. The splicing and cleavage reaction products are indicated by arrows.

and are summarized in Table 3. All of the experiments were performed in triplicate.

## RESULTS

**The *Spl* DnaX Mini-intein Is Catalytically Active.** To determine whether the *Spl* DnaX mini-intein is catalytically active, a His<sub>6</sub>-UBQ-Intein-HD-Flag<sub>3</sub> protein construct (Figure

2A and Figure S1), consisting of *Spl* DnaX mini-intein (INT) with ubiquitin (UBQ) and homeodomain (HD) as the N-extein and C-extein, respectively, was designed and expressed at 37 °C in *E. coli*. After induction, cells were collected at different time points, immediately lysed, and analyzed by SDS-PAGE followed by Western blotting (Figure 2B). The 37.6 kDa precursor protein was not detected at any time, indicating rapid



**Figure 3.** Expression and purification of *Spl* DnaX mini-intein. (A) 12% SDS–PAGE of intein<sup>WT</sup> (top) and intein<sup>C1A/C(+1)A</sup> (bottom) showing high protein purity. (B) Analytical SEC profile showing that both intein<sup>WT</sup> (top) and intein<sup>C1A/C(+1)A</sup> (bottom) eluted almost at the same volume. The dotted line is for protein standards. (C) SV-AUC statistical analysis of intein<sup>WT</sup> (top) and intein<sup>C1A/C(+1)A</sup> (bottom). The top panel in the figures shows the overlay of the experimental values (circle), and the fitted curve (solid line) with a root-mean-square deviation of 0.007 for both constructs. The bottom panel shows the corresponding residuals of the fitted data. (D) The plot of AUC distance distribution  $c(S)$  vs sedimentation coefficient (S) for intein<sup>WT</sup> (top) and intein<sup>C1A/C(+1)A</sup> (bottom) generated a single sharp peak indicating a monodispersed and homogeneous protein sample.

catalysis of the splicing reaction by the intein enzyme immediately upon synthesis in the cell. The ligated product (His<sub>6</sub>-UBQ-HD-Flag<sub>3</sub>, 21.9 kDa) and the excised intein (INT, 15.7 kDa) were detected, confirming that the *Spl* DnaX mini-intein is catalytically active. The precursor protein also undergoes side reactions where cleavage occurs at the N-terminus or the C-terminus of the intein (Figure 2A). C-Terminal cleavage products, i.e., His<sub>6</sub>-UBQ-INT (26.6 kDa) and HD-Flag<sub>3</sub> (11 kDa), were detected by Western blotting (Figure 2B). One of the N-terminal cleavage products, i.e., His<sub>6</sub>-UBQ (10.9 kDa), was also detected.

At 37 °C, the precursor protein was not detected at any time point of 10, 20, 40, or 60 min after induction, indicating rapid splicing and cleavage reactions upon the synthesis of the enzyme in the cell. To slow these catalytic reactions, the assay was repeated at 16 °C and both short and long time points were collected (Figure 2C). Again, the precursor protein was not detected, but the splicing and cleavage products were detected. The lack of detection of the precursor protein is not due to poor expression, as the spliced products have thick bands in both assays (Figure 2B,C). This rapid catalytic reaction also precluded purification of the precursor protein and any subsequent kinetics of the splicing reaction as has been performed for several other intein enzymes.<sup>2,49,50</sup> However, one can conclude that the *Spl* DnaX mini-intein is catalytically active and is one of the fastest intein enzymes.

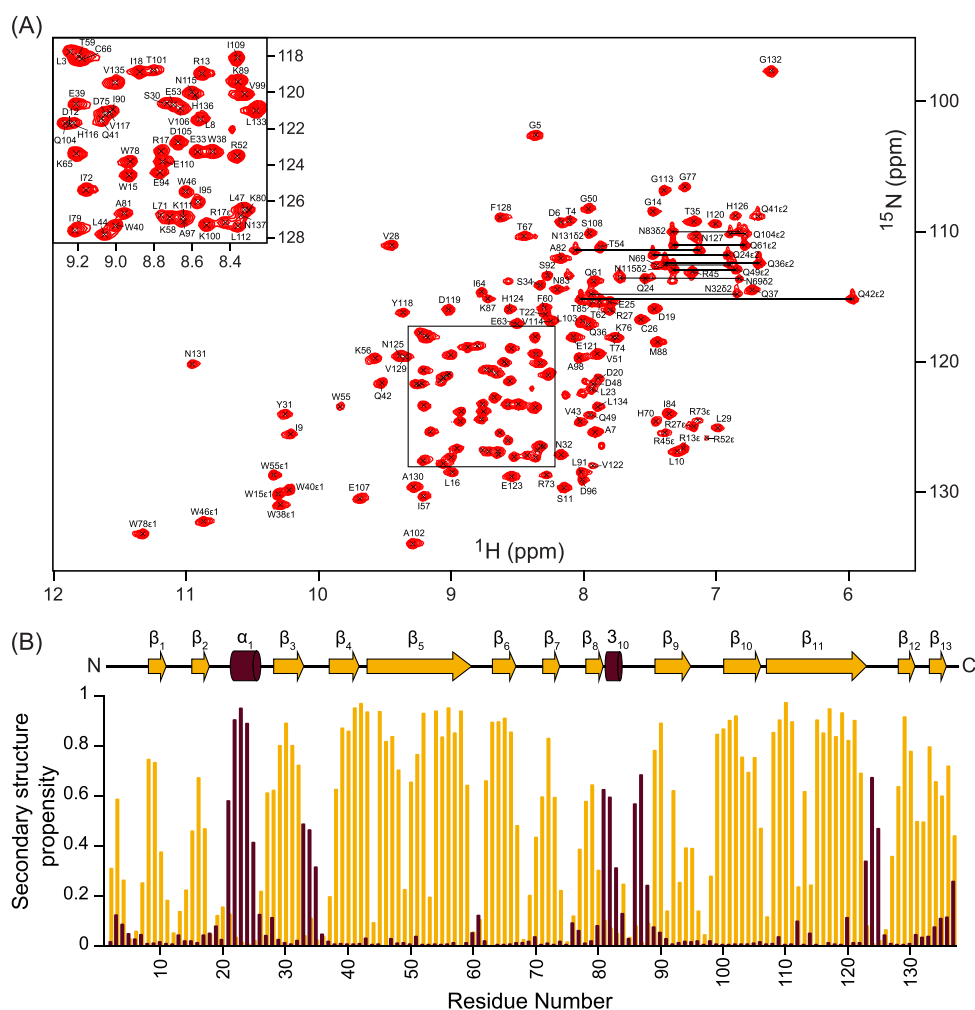
**The *Spl* DnaX Mini-intein Is a Monomer in Solution.** As the *Spl* DnaX mini-intein is highly active, the N-terminal His<sub>6</sub> tag is cleaved within the cell, thus hampering affinity purification of the protein and resulting in a poor yield of the purified protein.

Hence, residues C1 and C(+1) were mutated to alanines to enable high-quality purification with good yields, which is necessary for structural studies.

Both intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup> (Table 1) were obtained in high purity (Figure 3A) and were further characterized by analytical SEC and AUC. The SEC profiles confirm the high purity and monomeric state of both proteins (Figure 3B). Intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup> elute at 12.49 and 12.37 mL, respectively. Intein<sup>C1A/C(+1)A</sup> has five extra residues in the N-terminus compared to intein<sup>WT</sup> (Table 1), which most likely results in the slightly different elution volumes.

From the sedimentation velocity experiment of AUC, Stokes radii of 2.14 and 2.21 nm, sedimentation coefficients ( $S_w$ ) of 1.776 and 1.819 S, and frictional ratios ( $f/f_0$ ) of 1.26 and 1.28 were determined for intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup>, respectively (Figure 3C,D). The  $S_w$  values correlate to molecular weights of 16.9 and 17.7 kDa for intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup>, respectively, which closely match the expected molecular weights for the proteins. Additionally, the frictional ratio of ~1.3 strongly suggests that both are globular proteins.<sup>51</sup> The continuous  $c(s)$  distribution model for fitted data clearly indicates the presence of a monomeric protein as the dominant species in solution (Figure 3D). Overall, the SEC and AUC studies show that intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup> have similar hydrodynamic properties, indicating no perturbation due to mutation.

**Solution NMR Structure of *Spl* DnaX Mini-intein.** The structure of intein<sup>C1A/C(+1)A</sup> in solution was determined at pH 6.5 by NMR spectroscopy. The <sup>15</sup>N–<sup>1</sup>H HSQC spectrum of intein<sup>C1A/C(+1)A</sup> (Figure 4A) shows well-dispersed peaks indicating the presence of a properly folded domain. Similar



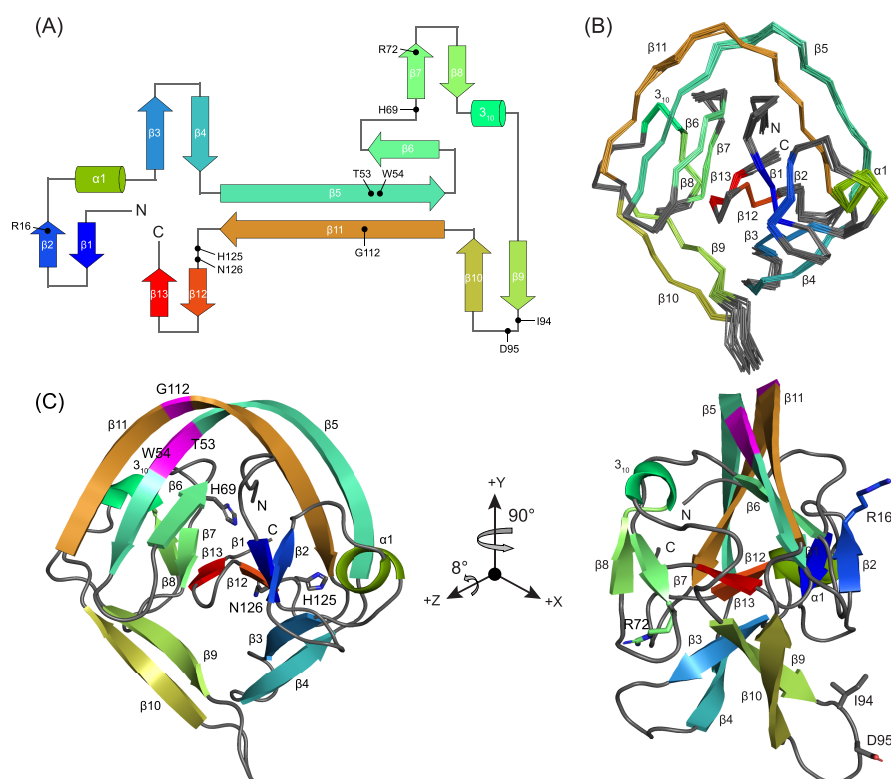
**Figure 4.** Backbone assignment and secondary structure prediction of *Spl* DnaX mini-intein. (A)  $^{15}\text{N}$ – $^1\text{H}$  HSQC spectrum of intein<sup>C1A/C(+1)A</sup> with the backbone assignments. The inset shows the labeled peaks in the central crowded region of the spectrum. Side chain resonances of Asn and Gln are connected with horizontal lines. (B) Secondary structure propensities of intein<sup>C1A/C(+1)A</sup> plotted for each residue. The propensities of strands and helices are colored orange and maroon, respectively. The resulting secondary structure is shown on the top with strands as arrows and helices as cylinders. The numbering starts with N-extein residue Glu(–1) as 1.

$^{15}\text{N}$ – $^1\text{H}$  HSQC spectra of intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup> indicate that the mutant maintains the same fold as the wild-type protein (Figure S2). The  $^{15}\text{N}$ – $^1\text{H}$  HSQC spectrum of intein<sup>C1A/C(+1)A</sup> was assigned using standard  $^{15}\text{N}$ ,  $^{13}\text{C}$ , and  $^1\text{H}$  heteronuclear NMR experiments. Intein<sup>C1A/C(+1)A</sup> contains 141 residues (Table 1). Nearly complete assignments of the backbone (99.85%) and the side chain (98%) chemical shifts (except the aromatic atoms) were obtained. The diastereotopic methyl groups present in leucines and valines were assigned by the fractional (10%)  $^{13}\text{C}$ -labeled sample.<sup>52</sup> Using the backbone  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$ , and  $^{13}\text{CO}$  chemical shifts, the secondary structure propensity was calculated with MICS.<sup>31</sup> It shows the presence of 13  $\beta$ -strands along with two helices for intein<sup>C1A/C(+1)A</sup> (Figure 4B). Intein<sup>C1A/C(+1)A</sup> has two cysteine residues C25 and C65 with  $^{13}\text{C}_\beta$  chemical shifts of 31.07 and 34.05 ppm, respectively. This indicates that C25 and C65 do not form a disulfide linkage between them.<sup>53</sup> Similarly, on the basis of  $\text{C}_\beta$  and  $\text{C}_\gamma$  chemical shifts,<sup>54</sup> all three proline residues in the protein were found to have *trans*-peptide bonds.

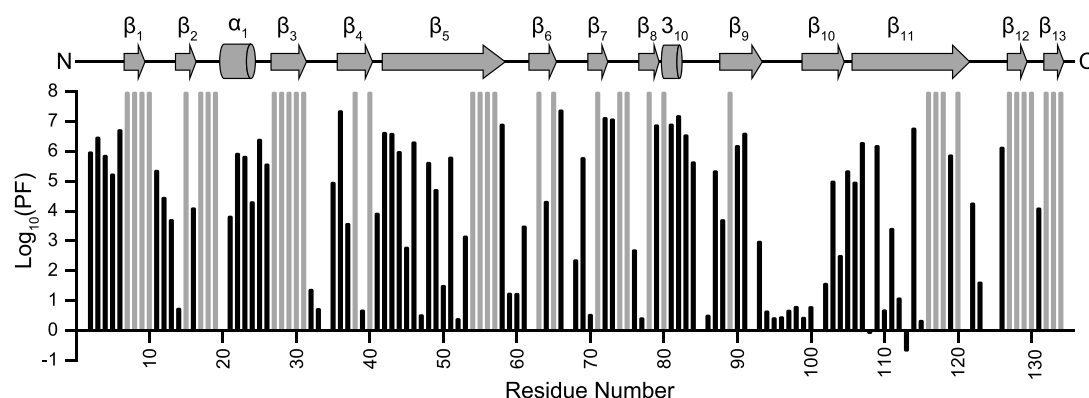
Distance and dihedral angle restraints were used to calculate the initial structures of the protein. Once consistent folded structures were obtained, hydrogen bond restraints, based on hydrogen exchange experiments (*vide infra*), were also added. In the final step, 100 structures were calculated, from which the statistics of the 15 best structures having the lowest energy are summarized in Table 2. On average, 17.5 restraints were obtained for each residue. Backbone  $\Phi$  and  $\Psi$  torsion angles of 97.1% residues are in favored regions of the Ramachandran plot.<sup>55</sup> None of the residues are in disallowed regions. These structures align with a root-mean-square deviation (RMSD) of 0.53 Å for all heavy atoms and 0.27 Å for backbone atoms across all ordered residues (Table 2). The elongated loop between strands  $\beta_9$  and  $\beta_{10}$  shows the highest degree of flexibility of this intein structure (Figure 5). NMR assignment data have been deposited in the BioMagResBank (entry 50361), and the structural coordinates have been deposited in the Protein Data Bank (PDB entry 7CFV).

**Amide Hydrogen Exchange Reveals a Highly Stable Core of *Spl* DnaX Mini-intein.** The residue-wise stability of





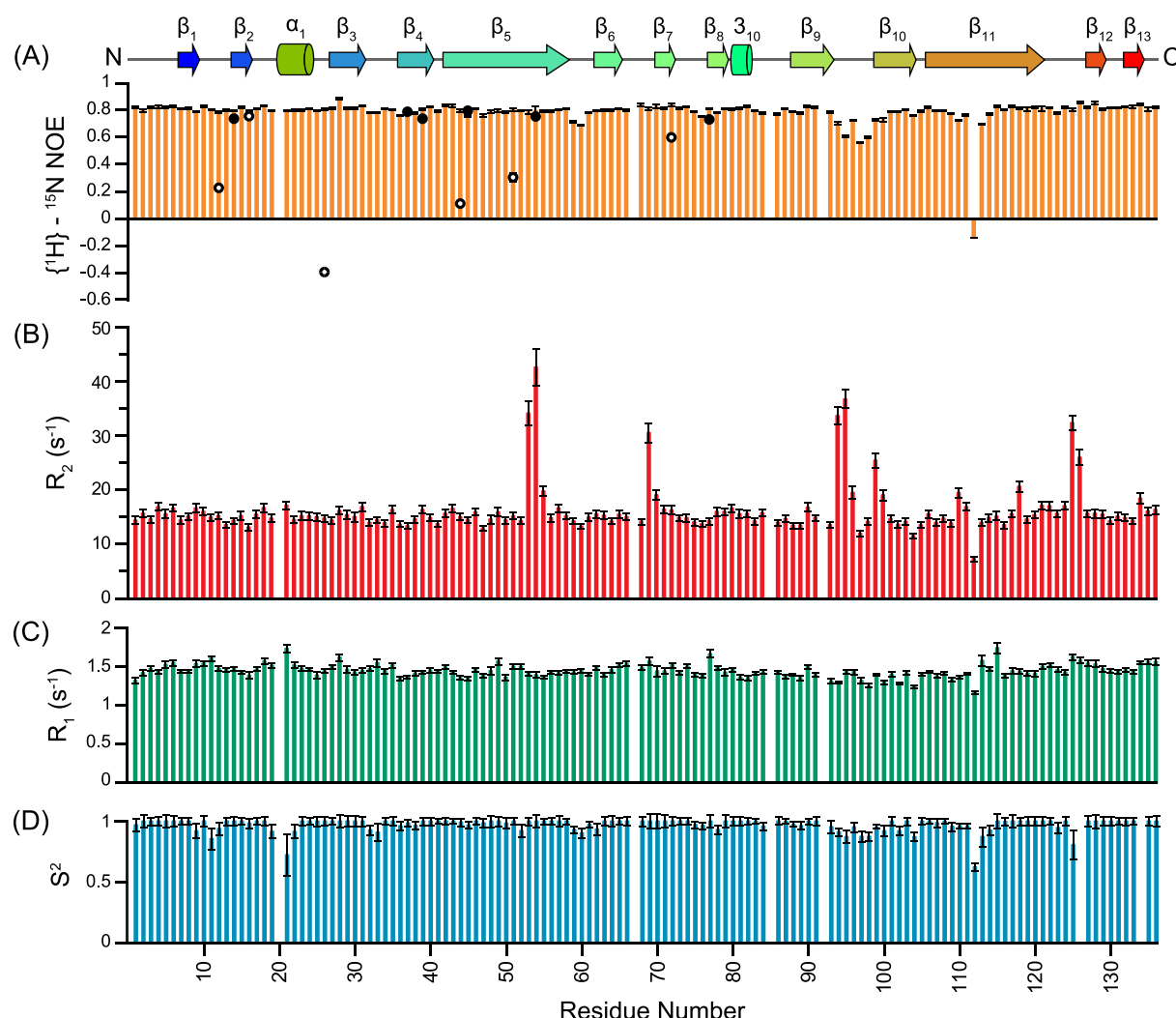
**Figure 5.** Structural ensembles of *Spl* DnaX mini-intein. (A) Secondary structure topology map of *Spl* DnaX mini-intein. The residues showing significant dynamic motions are labeled with filled circles. (B) Ensemble showing the  $C_{\alpha}$  trace of 15 energy-minimized conformers representing the three-dimensional structure of *Spl* DnaX mini-intein, with a backbone atom RMSD of 0.27 Å. Regions of the secondary structure are labeled. The N- and C-termini are indicated. (C) Ribbon representation of the lowest-energy conformer of *Spl* DnaX mini-intein. The dynamic residues are shown as sticks. The positions of residues T53, W54, and G112 are colored magenta.



**Figure 6.** Amide HX studies show that *Spl* DnaX mini-intein has a very stable core. Backbone amide HX studies demonstrate a wide range of protection factors (PFs). The slowest-exchanging amides, which show no appreciable decay over a period of 110 days, are assigned a protection factor (PF) of  $10^8$  and colored gray. The residues whose PFs were measured are colored black. The missing bars correspond to prolines or residues that do not fit well. The secondary structure is shown on the top with arrows and cylinders for strands and helices, respectively.

*Spl* DnaX mini-intein was determined by NMR-based hydrogen exchange (HX) experiments. Using complementary protium–deuterium HX (pH 6.5) and protium–protium CLEANEX-PM (pH 6.5 and 7.5) experiments, the protection factors (PFs) for the backbone amide protons were calculated (Figure 6). HX measurements of intein<sup>C1A/C(+1)A</sup> revealed a wide range of stability throughout the protein. Several amides did not display any significant decrease in peak intensity over a period of ~110 days, the course of the experiment. These amides have been assigned a PF of  $>10^8$  (Figure 6, gray bars). Their high protection indicates that most likely these amides exchange via a

global unfolding pathway. These amides are located mostly in the  $\beta$  strands and form the rigid hydrophobic core of the enzyme. Residues in the  $3_{10}$ -helix (PF  $\sim 10^7$ ),  $\alpha$ -helix (PF  $\sim 10^3$ – $10^6$ ), and the turn between the  $\alpha$ -helix and  $\beta_3$  (PF  $\sim 10^6$ – $10^7$ ) have intermediate protection, indicating subglobal fluctuations.<sup>43</sup> In contrast, amides in most loops have low protection factors (PF  $\sim 10$ – $10^4$ ) and most likely exchange via local conformational fluctuations. In particular, the residues in the loop between  $\beta_9$  and  $\beta_{10}$  have very low protection factors (PF  $< 10$ ). This is consistent with its solvent-exposed structure (Figure 5C).



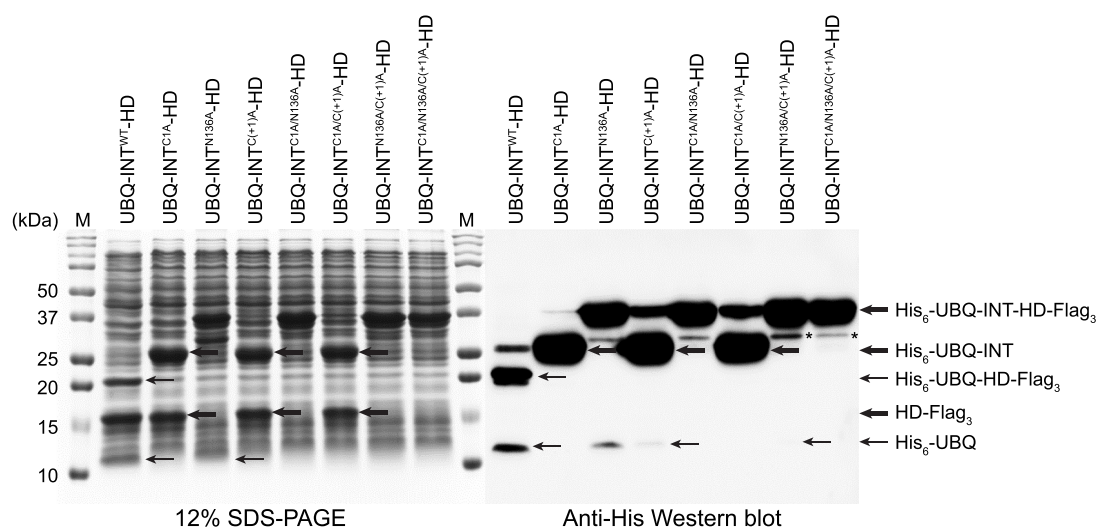
**Figure 7.** Fast nanosecond to picosecond time scale mobility of amides in *Spl* DnaX mini-intein. (A) Heteronuclear  $\{^1\text{H}\}-^{15}\text{N}$  NOEs, (B) transverse relaxation rate constants ( $R_2$ ), (C) longitudinal relaxation rate constants ( $R_1$ ), and (D) order parameters ( $S^2$ ) plotted for each residue. The  $\{^1\text{H}\}-^{15}\text{N}$  NOE values of tryptophan (filled circles) and arginine (empty circles) side chains are also indicated. The secondary structure is shown on the top with arrows and cylinders for strands and helices, respectively.

**$^{15}\text{N}$  Relaxation Measurements Reveal Fast Dynamics of the *Spl* DnaX Mini-intein.** Amide  $^{15}\text{N}$  relaxation ( $R_1$ ,  $R_2$ , and steady-state heteronuclear NOE) experiments were performed at 30 °C for intein<sup>C1A/C(+1)A</sup> (Figure 7). For almost all residues, the heteronuclear  $\{^1\text{H}\}-^{15}\text{N}$  NOE values are in the range of 0.55–0.88 (Figure 7A), indicating the presence of a well-folded structure.<sup>41</sup> The residues in the loop between  $\beta_9$  and  $\beta_{10}$  have low  $\{^1\text{H}\}-^{15}\text{N}$  NOE values, consistent with their low protection factors, and indicate local flexibility. The side chain  $\text{N}_{\epsilon 1}-\text{H}_{\epsilon 1}$  pairs of all six tryptophans have  $\{^1\text{H}\}-^{15}\text{N}$  NOE values of  $>0.7$ , indicating that these side chains have highly restricted mobility (Figure 7A). This is consistent with their buried conformation in the intein structure. Similarly, of six arginine residues, the side chain  $\text{N}_{\epsilon}-\text{H}_{\epsilon}$  pairs of R16 and R72 have  $\{^1\text{H}\}-^{15}\text{N}$  NOE values of 0.75 and 0.6, respectively, indicating the rigid nature of these long arginine side chains (Figure 7A). The R16 and R72 side chains are within 4 Å of D5 and E32 side chains, respectively. These salt-bridge interactions most likely result in the rigidity of these arginine side chains and should also contribute to the stability of the intein structure.

Transverse relaxation rate constants ( $R_2$ ) have a value of  $\sim 15 \text{ s}^{-1}$  for most of the residues (Figure 7B). Some residues have

significantly higher  $R_2$  values, indicating the presence of slower microsecond to millisecond time scale motions. Most of these residues are present in the loop regions. I94 and D95 are present in the loop between strands  $\beta_9$  and  $\beta_{10}$ ; H125 and D126 are present in the loop between strands  $\beta_{11}$  and  $\beta_{12}$ , and His69 is present in the conserved TXXH motif between strands  $\beta_6$  and  $\beta_7$  (Figure 5). Interestingly, residues T53 and W54, which are present at the center of strand  $\beta_5$ , also have high  $R_2$  values. The longitudinal relaxation rate constants ( $R_1$ ) have much less variation with an average value of  $\sim 1.4 \text{ s}^{-1}$  (Figure 7C). The residue-specific order parameter ( $S^2$ ), which gives a measure of the residue-wise rigidity of the protein, was calculated using TENSOR2.<sup>42</sup> Most of the residues, including the N- and C-termini, exhibit high order parameters. The average order parameter of the protein is  $0.97 \pm 0.03$  (Figure 7D). Using the  $R_2/R_1$  ratio, the isotropic tumbling correlation time,  $\tau_c$ , for intein<sup>C1A/C(+1)A</sup> was calculated to be  $9.84 \pm 0.03 \text{ ns}$ , indicating its consistency with the molecular mass of the protein and its monomeric state.

**Identification of the Active Site Residues and Their Roles in Catalysis.** The three-dimensional structure of *Spl* DnaX mini-intein and previous work on other intein



**Figure 8.** Comparative assay of *Spl* DnaX intein mutants. On the left, 12% SDS–PAGE analysis of the wild type and mutants of *Spl* DnaX mini-intein expressed at 37 °C. The cells of all of the constructs were harvested 1 h after induction. On the right, Western blot using the anti-His<sub>6</sub> antibody showing the spliced and side reaction cleavage products, as indicated by arrows. The asterisks denote the nonspecific degradation products.

enzymes<sup>13,56–60</sup> suggested that residues C1, N136, and C(+1) may be directly involved in the catalysis of the splicing reaction (Figure 1). To determine their roles in the catalysis of splicing as well as the N- and C-terminal cleavage reactions, a series of mutations were designed where these three residues were mutated to alanines in all possible combinations of single, double, and triple mutations (Table 1). The double and triple mutants were designed to study whether the catalytic functions of these active site residues were coupled to each other. The effect of these mutations on the in cell splicing after induction for 1 h was monitored by SDS–PAGE. The splicing and the N- and C-terminal cleavage products were specifically detected by Western blot using an anti-His<sub>6</sub> antibody and are summarized in Table 3.

None of the mutants resulted in the spliced product, which indicates that these residues are indispensable for the protein splicing reaction catalyzed by *Spl* DnaX mini-intein (Figure 8). The substitution of C1 with alanine completely blocked the splicing and N-terminal cleavage reactions in all four C1A mutants. The C-terminal cleavage products, His<sub>6</sub>-UBQ-INT (26.6 kDa) and HD-Flag<sub>3</sub> (11 kDa), were detected for UBQ-INT<sup>C1A</sup>-HD and UBQ-INT<sup>C1A/C(+1)A</sup>-HD. Thus, C1 is necessary for the initial nucleophilic attack on the preceding peptide bond and the formation of the linear thioester intermediate (Figure 1, step 1). The C(+1) to alanine mutation [UBQ-INT<sup>C(+1)A</sup>-HD] prevents the splicing reaction but results in the cleavage at both N- and C-termini of the intein enzyme. Thus, C(+1) is necessary for the formation of the branched thioester intermediate (Figure 1, step 2).

Mutation of N136 to alanine prevented the splicing and C-terminal cleavage reactions in all four N136A mutants. However, the N-terminal cleavage product His<sub>6</sub>-UBQ was observed for UBQ-INT<sup>N136A</sup>-HD and UBQ-INT<sup>N136A/C(+1)A</sup>-HD (Figure 8). Thus, the C-terminal cleavage reaction is catalyzed by Asn136 (Figure 1, step 3). The triple mutant UBQ-INT<sup>C1A/N136A/C(+1)A</sup>-HD yielded the stable precursor protein with a molecular weight of 37.6 kDa, which neither spliced nor cleaved at the two termini. Thus, C1, N136, and C(+1) are the active site residues, which are directly involved in the catalysis of the splicing reaction. Similar <sup>15</sup>N–<sup>1</sup>H HSQC spectra of intein<sup>WT</sup> and intein<sup>C1A/C(+1)A</sup> show that the two proteins maintain the same structural fold

(Figure S2). This confirms that the loss of function is not due to improper folding of these mutants.

## DISCUSSION

***Spl* DnaX Mini-intein Has a Conserved HINT Domain Structure.** The *Spl* DnaX mini-intein is one of the smallest intein enzymes (136 residues) characterized so far (Table S1). Its sequence was compared with those of other inteins, whose structure has been determined (Figure S3), and the highest level of sequence identity of 32% was obtained with *Spl* DnaB intein (Table S1). Although the level of sequence identity is low, *Spl* DnaX mini-intein has very good structural homology with other inteins (Table S1). The *Spl* DnaX mini-intein comprises 13  $\beta$ -strands, one  $\alpha$ -helix, and one  $3_{10}$ -helix, which are arranged in a compact horseshoe-shaped fold commonly found in the HINT (hedgehog/intein) domain superfamily (Figure 5). The  $3_{10}$ -helical turn was confirmed by observing a potential hydrogen bond distance of  $\sim 2.4$  Å in all of the models between the carbonyl oxygen of A80 and the amide proton of I83, thus involving 10 backbone atoms in a turn. In inteins, the endonuclease domain is typically inserted within the HINT fold between strands  $\beta$ 10 and  $\beta$ 11. In *Spl* DnaX mini-intein, the endonuclease domain is replaced by an extended loop between strands  $\beta$ 9 and  $\beta$ 10. This loop does not affect the overall HINT fold of the *Spl* DnaX mini-intein. Thus, the structure of *Spl* DnaX mini-intein is highly similar with other intein structures.<sup>21,61–65</sup>

On the basis of the determined structure of *Spl* DnaX mini-intein and the sequence alignment with other inteins (Figure S3), we have identified the four conserved blocks of residues in *Spl* DnaX mini-intein, i.e., blocks A, B, F, and G, which are characteristic of all inteins.<sup>66,67</sup> Block A begins with the conserved Cys nucleophile (C1) and consists of 13 residues. Usually, 60–100 residues from the N-terminus, block B is present and contains a conserved TXXH motif. Residues <sup>66</sup>TGNH<sup>69</sup> form this motif in *Spl* DnaX mini-intein. Blocks C–E and H are present in the endonuclease domain and, hence, not observed in *Spl* DnaX mini-intein. Block F in inteins contains a conserved Asp and a conserved HNF motif, which are D118 and <sup>125</sup>HNF<sup>127</sup>, respectively, in *Spl* DnaX mini-intein. Lastly, block G



ends with a conserved HNC motif, which are residues H135–N136–C(+1) in *Spl* DnaX mini-intein.

**Symmetry in Structure and Dynamic Motions in *Spl* DnaX Mini-intein.** The *Spl* DnaX mini-intein has a highly symmetric structure.<sup>68</sup> The two large  $\beta$ -strands,  $\beta 5$  and  $\beta 11$ , form a long, curving antiparallel  $\beta$ -sheet that results in the horseshoe shape of the intein enzymes. The extreme curvature of these strands puts a high strain on this structure, which is relieved by dynamic motions. In strand  $\beta 5$ , the backbone amides of residues T53 and W54 have significantly higher  $R_2$  relaxation rate constants (Figure 7B), indicating motions in the slow microseconds to millisecond time scale, characteristic of conformational exchange.<sup>69</sup> On the contrary, G112 in strand  $\beta 11$  exhibits faster motion on the picosecond to nanosecond time scale as demonstrated by its  $\{^1\text{H}\}$ – $^{15}\text{N}$  NOE value of  $-0.14$  (Figure 7A). Interestingly, these residues in strands  $\beta 5$  and  $\beta 11$  are close in space and cluster together (Figure 5C). The HX experiments also reveal that several residues in strands  $\beta 5$  and  $\beta 11$  have relatively less protection from the rest of the protein (Figure 6). Similar dynamic motions and low protection factors in the corresponding  $\beta$ -strands have also been observed for *Mtu* RecA<sup>22</sup> and *Pab* PolII<sup>23</sup> inteins and might be a conserved feature of the intein structure.

The *Spl* DnaX mini-intein has several  $\beta$ -hairpins. The  $\beta 1$ – $\beta 2$  and  $\beta 7$ – $\beta 8$   $\beta$ -hairpin pairs as well as the  $\beta 3$ – $\beta 4$  and  $\beta 9$ – $\beta 10$   $\beta$ -hairpin pairs are symmetrically positioned in the structure (Figure 5). The dynamically rigid arginines R16 and R72 (Figure 7A) occupy symmetrically equivalent positions on strands  $\beta 2$  and  $\beta 7$  (Figure 5). Moreover, several loop residues such as H69, present in the conserved TXXH motif, I94 and D95, present in the loop between strands  $\beta 9$  and  $\beta 10$ , and H125 and N126, present in the loop between strands  $\beta 11$  and  $\beta 12$ , have significantly higher  $R_2$  relaxation rate constants (Figure 7B). Thus, these residues also exhibit slow microsecond to millisecond time scale motions resulting from a conformational exchange. Similar dynamic motions in the corresponding regions have been observed for *Mja* KlbA,<sup>20</sup> *Npu* DnaE,<sup>21</sup> *Pho* RadA,<sup>70</sup> and *Mtu* RecA<sup>22</sup> inteins. Thus, inteins appear to have conserved motions, some of which are in structurally symmetric regions, which most likely is a consequence of the symmetric intein structure.

**Mechanism of the Splicing Reaction Mediated by *Spl* DnaX Mini-intein.** On the basis of their reaction mechanisms (Figure 1), intein enzymes have been categorized into three classes.<sup>12</sup> The *Spl* DnaX mini-intein has four cysteines, C1, C25, C65, and C(+1). Mutation of C1 and C(+1), alone or in combination, completely abolished the splicing reaction (Figure 8). Thus, the two-terminal cysteines, and not the intermediary C25 and C65, are involved in catalyzing the splicing reaction, confirming that *Spl* DnaX mini-intein belongs to the type I class of inteins.<sup>12</sup> The C1A mutation in all four mutants prevented the nucleophilic attack of the thiol group of C1 on the carbonyl group of the previous peptide bond between E(–1) and C1 and, thus, blocked the N–S acyl rearrangement, which prevents the formation of the linear thioester intermediate (Figure 1A, step 1). Attack on the linear thioester intermediate by the C(+1) thiol results in the formation of the branched thioester intermediate and subsequently the spliced product (Figure 1A, step 2). Hydrolysis of the thioester intermediates results in N-terminal cleavage. Hence, the C1A mutation prevented both the splicing and N-terminal cleavage reactions (Table 3).

The C(+1) thiol attacks the linear thioester intermediate and forms the branched intermediate (Figure 1, step 2). Both UBQ-

INT<sup>N136A</sup>-HD and UBQ-INT<sup>N136A/C(+1)A</sup>-HD mutants are incapable of splicing and C-terminal cleavage reactions due to the absence of N136. However, UBQ-INT<sup>N136A</sup>-HD can form both linear and branched intermediates, whereas UBQ-INT<sup>N136A/C(+1)A</sup>-HD can form only the linear intermediate. Hydrolysis of these intermediates results in the N-terminal cleavage products. As UBQ-INT<sup>N136A</sup>-HD can form both intermediates, this results in a larger amount of N-terminal cleavage product relative to UBQ-INT<sup>N136A/C(+1)A</sup>-HD (Table 3). Consistent with this, the UBQ-INT<sup>C(+1)A</sup>-HD mutant undergoes relatively less N-terminal cleavage compared to UBQ-INT<sup>N136A</sup>-HD (Table 3).

The N136A mutation prevents the splicing and C-terminal cleavage reactions, indicating that N136 is necessary for the C-terminal cleavage step (Figure 1A, step 3). Interestingly, the level of the C-terminal cleavage product is consistently low for the UBQ-INT<sup>C(+1)A</sup>-HD and UBQ-INT<sup>C1A/C(+1)A</sup>-HD mutants compared to that for the UBQ-INT<sup>C1A</sup>-HD mutant (Table 3). This indicates that C(+1) most likely facilitates the Asn cyclization and subsequent C-terminal cleavage (Figure 1A, step 3). Collectively, these mutagenesis experiments define the roles of the three conserved splice junction residues.

**Competition between Splicing and Cleavage Reactions in *Spl* DnaX Mini-intein.** Apart from the splicing reaction, inteins also undergo N- and C-terminal cleavage reactions. Because inteins are single-turnover enzymes, these reactions decrease the overall yield of the spliced product. We designed all possible alanine mutations of the active site residues C1, N136, and C(+1) of *Spl* DnaX intein and functionally characterized them. These studies also provided important insights into the competition between the splicing and cleavage reactions and the underlying roles of the active site residues.

Splicing occurred only in the wild-type enzyme and requires all three active site residues. Among the cleavage reactions, the C-terminal cleavage is highly efficient, which is evident from the single mutants (Table 3). For the N-terminal cleavage incompetent mutant UBQ-INT<sup>C1A</sup>-HD, 90% C-terminal cleavage was detected, whereas for the C-terminal cleavage incompetent mutant UBQ-INT<sup>N136A</sup>-HD, only 7.5% N-terminal cleavage was detected. Thus, it appears that the N- and C-terminal cleavage reactions can proceed independently in *Spl* DnaX intein mutants. This is consistent with similar observations for *Ssp* DnaB intein.<sup>61</sup> It is important to note that according to the class 1 reaction mechanism (Figure 1), the C-terminal cleavage should occur after the N-terminal nucleophilic attack and the formation of the branched intermediate to form the spliced product. A precursor protein undergoing C-terminal cleavage before branched intermediate formation will not produce the spliced product. Hence, the suppression of premature C-terminal cleavage is necessary for a successful splicing reaction. In the wild-type *Spl* DnaX intein, the spliced product is indeed the major product ( $\sim 60\%$ ). Interestingly, the N- and C-terminal cleavage products are almost equal ( $\sim 20\%$ ) for the wild-type intein. This indicates that either cysteines C1 and C(+1) suppress the reactivity of N136 or N136 enhances the reactivity of the cysteines in the wild-type intein, and thus, favors the formation of the spliced product. Overall, this highlights the interdependencies of the catalytic residues in driving the splicing reaction.

## CONCLUSION

In summary, we have determined the solution NMR structure of a new intein sequence derived from the cyanobacterium *S.*

*platensis*. We show that the *Spl* DnaX mini-intein is monomeric in solution and has a typical HINT domain fold. We also characterized the dynamics of the protein and found conserved motions in the protein. NMR-based hydrogen exchange experiments revealed the presence of a highly stable core in the *Spl* DnaX mini-intein. In cell splicing assays demonstrated that *Spl* DnaX mini-intein is a highly active enzyme. We further investigated the functional roles of the catalytic residues by designing combinatorial alanine mutations, which showed their combined effects in suppressing N- and C-terminal cleavage reactions and enhancing the splicing product. The small size, high stability, and high catalytic activity make the *Spl* DnaX intein enzyme an ideal candidate for protein engineering that may lead to novel applications in protein biochemistry.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.0c00828>.

Comparison of *Spl* DnaX mini-intein with other intein structures using the DALI server (Table S1), complete sequence of the fusion construct (Figure S1), similar fold in wild-type and mutant inteins (Figure S2), and structure-based sequence alignment of the *Spl* DnaX mini-intein with other inteins (Figure S3) (PDF)

### Accession Codes

UniprotKB, K1XG21; PDB, 7CFV; BMRB, 50361.

## ■ AUTHOR INFORMATION

### Corresponding Author

**Soumya De** – School of Bioscience, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India; [orcid.org/0000-0002-4347-3137](https://orcid.org/0000-0002-4347-3137); Phone: (03222) 260514; Email: [somde@iitkgp.ac.in](mailto:somde@iitkgp.ac.in)

### Authors

**Soumendu Boral** – School of Bioscience, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India

**Snigdha Maiti** – School of Bioscience, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India

**Aditya J. Basak** – School of Bioscience, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India

**Woonghee Lee** – Department of Chemistry, University of Colorado Denver, Denver, Colorado 80217, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.biochem.0c00828>

### Author Contributions

S.B. performed the experiments. S.B., A.J.B., W.L., and S.D. processed, analyzed, and determined the NMR structure. S.B., S.M., and S.D. processed and analyzed the dynamics data. S.B. and S.D. conceived the study and designed experiments. All authors wrote and revised the manuscript.

### Funding

S.B. acknowledges a NET-JRF fellowship [918/(CSIR-UGC NET DEC. 2016)] provided by the University Grants Commission (UGC), Government of India. S.M. and A.J.B. acknowledge IIT Kharagpur for a fellowship. This work has been supported by the Science and Engineering Research Board

(SERB), an India Early Career Research (ECR) award (ECR/2016/000847, dated 07/03/2017), and an ISIRD grant from IIT Kharagpur (IIT/SRIC/ISIRD/2015–2016, dated 14/12/2015) to S.D. W.L. and computational resources at NMRFAM have been supported by National Science Foundation Grant DBI 1902076, UC-Denver STYPE-61193205, and National Institutes of Health Grant P41GM103399 (National Institutes of General Medical Sciences).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors acknowledge Dr. Sayantani Roy and Saptarshi Banerjee for helpful discussions. The authors thank the Central Research Facility (CRF) at IIT Kharagpur for the use of NMR, SEC, and AUC facilities.

## ■ ABBREVIATIONS

NMR, nuclear magnetic resonance; HINT, hedgehog intein; *Spl*, *S. platensis*; IPTG, isopropyl  $\beta$ -thiogalactopyranoside; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; BME,  $\beta$ -mercaptoethanol; PMSF, phenylmethanesulfonyl fluoride; HSQC, heteronuclear single-quantum coherence; PDB, Protein Data Bank; BMRB, Biological Magnetic Resonance Data Bank; PVDF, polyvinylidene difluoride; TBST, Tris-buffered saline with Tween 20.

## ■ REFERENCES

- (1) Kane, P., Yamashiro, C., Wolczyk, D., Neff, N., Goebel, M., and Stevens, T. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 250, 651–657.
- (2) Xu, M.-Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Perler, F. B. (1993) In vitro protein splicing of purified precursor and the identification of a branched intermediate. *Cell* 75, 1371–1377.
- (3) Paulus, H. (2000) Protein Splicing and Related Forms of Protein Autoprocessing. *Annu. Rev. Biochem.* 69, 447–496.
- (4) Volkmann, G., and Mootz, H. D. (2013) Recent progress in intein research: from mechanism to directed evolution and applications. *Cell. Mol. Life Sci.* 70, 1185–1206.
- (5) Wood, D. W., and Camarero, J. A. (2014) Intein Applications: From Protein Purification and Labeling to Metabolic Control Methods. *J. Biol. Chem.* 289, 14512–14519.
- (6) Muir, T. W., Sondhi, D., and Cole, P. A. (1998) Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6705–6710.
- (7) Mootz, H. D., Blum, E. S., Tyszkiewicz, A. B., and Muir, T. W. (2003) Conditional Protein Splicing: A New Tool to Control Protein Structure and Function in Vitro and in Vivo. *J. Am. Chem. Soc.* 125, 10561–10569.
- (8) Kurpiers, T., and Mootz, H. D. (2008) Site-Specific Chemical Modification of Proteins with a Prelabeled Cysteine Tag Using the Artificially Split Mxe GyrA Intein. *ChemBioChem* 9, 2317–2325.
- (9) Liu, D., Xu, R., and Cowburn, D. (2009) Segmental Isotopic Labeling of Proteins for Nuclear Magnetic Resonance. In *Methods in Enzymology*, 1st ed., pp 151–175, Elsevier Inc.
- (10) Evans, T. C., Benner, J., and Xu, M. (1999) The Cyclization and Polymerization of Bacterially Expressed Proteins Using Modified Self-splicing Inteins. *J. Biol. Chem.* 274, 18359–18363.
- (11) Perler, F. B. (2002) InBase: the Intein Database. *Nucleic Acids Res.* 30, 383–384.
- (12) 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. *J. Biol. Chem.* 285, 2515–2526.

- (13) Xu, M. Q., and Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J.* 15, 5146–5153.
- (14) 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.
- (15) 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.
- (16) Duan, X., Gimble, F. S., and Quijcho, F. A. (1997) Crystal Structure of PI-SceI, a Homing Endonuclease with Protein Splicing Activity. *Cell* 89, 555–564.
- (17) Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., 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.
- (18) Poland, B. W., Xu, M., and Quijcho, F. A. (2000) Structural Insights into the Protein Splicing Mechanism of PI-SceI. *J. Biol. Chem.* 275, 16408–16413.
- (19) Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y., and Satow, Y. (2002) Protein-splicing Reaction via a Thiazolidine Intermediate: Crystal Structure of the VMA1-derived Endonuclease Bearing the N and C-terminal Propeptides. *J. Mol. Biol.* 316, 919–929.
- (20) Johnson, M. A., Southworth, M. W., Herrmann, T., Brace, L., Perler, F. B., and Wüthrich, K. (2007) NMR structure of a KlbA intein precursor from *Methanococcus jannaschii*. *Protein Sci.* 16, 1316–1328.
- (21) Oemig, J. S., Aranko, A. S., Djupsjöbacka, J., Heinämäki, 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.
- (22) Du, Z., Liu, Y., Ban, D., Lopez, M. M., Belfort, M., and Wang, C. (2010) Backbone Dynamics and Global Effects of an Activating Mutation in Minimized Mtu RecA Intein. *J. Mol. Biol.* 400, 755–767.
- (23) Du, Z., Liu, J., Albracht, C. D., Hsu, A., Chen, W., Marieni, M. D., Colelli, K. M., Williams, J. E., Reitter, J. N., Mills, K. V., and Wang, C. (2011) Structural and Mutational Studies of a Hyperthermophilic Intein from DNA Polymerase II of *Pyrococcus abyssi*. *J. Biol. Chem.* 286, 38638–38648.
- (24) Goldberg, R. J. (1953) Sedimentation in the Ultracentrifuge. *J. Phys. Chem.* 57, 194–202.
- (25) Schuck, P. (2000) Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* 78, 1606–1619.
- (26) Brautigam, C. A. (2015) Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. In *Analytical Ultracentrifugation*, 1st ed., pp 109–133, Elsevier Inc.
- (27) Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31, 1325–1327.
- (28) Sattler, M. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93–158.
- (29) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–93.
- (30) Lee, W., Westler, W. M., Bahrami, A., Eghbalnia, H. R., and Markley, J. I. (2009) PINE-SPARKY: Graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics* 25, 2085–2087.
- (31) Shen, Y., and Bax, A. (2012) Identification of helix capping and  $\beta$ -turn motifs from NMR chemical shifts. *J. Biomol. NMR* 52, 211–232.
- (32) Lee, W., Kim, J. H., Westler, W. M., and Markley, J. L. (2011) PONDEROSA, an automated 3D-NOESY peak picking program, enables automated protein structure determination. *Bioinformatics* 27, 1727–1728.
- (33) Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Marius Clore, G. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 160, 65–73.
- (34) Lee, W., Petit, C. M., Cornilescu, G., Stark, J. L., and Markley, J. L. (2016) The AUDANA algorithm for automated protein 3D structure determination from NMR NOE data. *J. Biomol. NMR* 65, 51–57.
- (35) Lee, W., Stark, J. L., and Markley, J. L. (2014) PONDEROSA-C/S: client-server based software package for automated protein 3D structure determination. *J. Biomol. NMR* 60, 73–75.
- (36) Lee, W., Cornilescu, G., Dashti, H., Eghbalnia, H. R., Tonelli, M., Westler, W. M., Butcher, S. E., Henzler-Wildman, K. A., and Markley, J. L. (2016) Integrative NMR for biomolecular research. *J. Biomol. NMR* 64, 307–332.
- (37) Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* 10, 980–980.
- (38) Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins: Struct., Funct., Genet.* 66, 778–795.
- (39) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
- (40) De, S., Chan, A. C. K., Coyne, H. J., Bhachech, N., Hermsdorf, U., Okon, M., Murphy, M. E. P., Graves, B. J., and McIntosh, L. P. (2014) Steric Mechanism of Auto-Inhibitory Regulation of Specific and Non-Specific DNA Binding by the ETS Transcriptional Repressor ETV6. *J. Mol. Biol.* 426, 1390–1406.
- (41) Maiti, S., Acharya, B., Boorla, V. S., Manna, B., Ghosh, A., and De, S. (2019) Dynamic Studies on Intrinsically Disordered Regions of Two Paralogous Transcription Factors Reveal Rigid Segments with Important Biological Functions. *J. Mol. Biol.* 431, 1353–1369.
- (42) Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* 16, 23–28.
- (43) Coyne, H. J., De, S., Okon, M., Green, S. M., Bhachech, N., Graves, B. J., and McIntosh, L. P. (2012) Autoinhibition of ETV6 (TEL) DNA Binding: Appended Helices Sterically Block the ETS Domain. *J. Mol. Biol.* 421, 67–84.
- (44) De, S., Okon, M., Graves, B. J., and McIntosh, L. P. (2016) Autoinhibition of ETV6 DNA Binding Is Established by the Stability of Its Inhibitory Helix. *J. Mol. Biol.* 428, 1515–1530.
- (45) Hwang, T. L., Van Zijl, P. C. M., and Mori, S. (1998) Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* 11, 221–226.
- (46) Zhang, Y.-Z. (1995) Protein and peptide structure and interactions studied by hydrogen exchange and NMR. Ph.D. Thesis, University of Pennsylvania, Philadelphia.
- (47) Connelly, G. P., and McIntosh, L. P. (1998) Characterization of a Buried Neutral Histidine in *Bacillus circulans* Xylanase: Internal Dynamics and Interaction with a Bound Water Molecule. *Biochemistry* 37, 1810–1818.
- (48) Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.
- (49) 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.
- (50) Frutos, S., Goger, M., Giovani, B., Cowburn, D., and Muir, T. W. (2010) Branched intermediate formation stimulates peptide bond cleavage in protein splicing. *Nat. Chem. Biol.* 6, 527–533.
- (51) Erickson, H. P. (2009) Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proced. Online* 11, 32–51.
- (52) Neri, D., Szyperski, T., Otting, G., Senn, H., and Wüthrich, K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. *Biochemistry* 28, 7510–7516.



- (53) Sharma, D., and Rajarathnam, K. (2000)  $^{13}\text{C}$  NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* 18, 165–71.
- (54) Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on  $^{13}\text{C}$  chemical shift statistics. *J. Biomol. NMR* 24, 149–54.
- (55) Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99.
- (56) Paulus, H. (2001) Inteins as Enzymes. *Bioorg. Chem.* 29, 119–129.
- (57) Evans, T. C., and Xu, M. (2002) Mechanistic and Kinetic Considerations of Protein Splicing. *Chem. Rev.* 102, 4869–4884.
- (58) Saleh, L., and Perler, F. B. (2006) Protein splicing in cis and in trans. *Chem. Rev.* 6, 183–193.
- (59) Mills, K. V., and Paulus, H. (2005) Biochemical Mechanisms of Intein-Mediated Protein Splicing. In *Homing Endonucleases and Inteins* (Belfort, M., et al., Eds.) pp 233–255, Springer-Verlag, Berlin.
- (60) Perler, F. (2005) Protein Splicing Mechanisms and Applications. *IUBMB Life* 57, 469–476.
- (61) Ding, Y., Xu, M., 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.
- (62) Aranko, A. S., Oeemig, J. S., Zhou, D., Kajander, T., Wlodawer, A., and Iwai, H. (2014) Structure-based engineering and comparison of novel split inteins for protein ligation. *Mol. Biosyst.* 10, 1023–1034.
- (63) Sun, P., Ye, S., Ferrandon, S., Evans, T. C., Xu, M., 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.
- (64) Beyer, H. M., Mikula, K. M., Li, M., Wlodawer, A., and Iwai, H. (2020) The crystal structure of the naturally split gp41–1 intein guides the engineering of orthogonal split inteins from cis-splicing inteins. *FEBS J.* 287, 1886–1898.
- (65) 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.
- (66) 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.
- (67) Pietrokovski, S. (1998) Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci.* 7, 64–71.
- (68) Hall, T. M. T., 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.
- (69) Palmer, A. G., Kroenke, C. D., and Patrick Loria, J. (2001) Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. *Methods Enzymol.* 339, 204–238.
- (70) Oeemig, J. S., Zhou, D., Kajander, T., Wlodawer, A., and Iwai, H. (2012) NMR and Crystal Structures of the *Pyrococcus horikoshii* RadA Intein Guide a Strategy for Engineering a Highly Efficient and Promiscuous Intein. *J. Mol. Biol.* 421, 85–99.