Structural and Mutational Studies of a Hyperthermophilic Intein from DNA Polymerase II of *Pyrococcus abyssi* 

**Received for publication, August 8, 2011, and in revised form, September 9, 2011**  
**Published, JBC Papers in Press, September 13, 2011, DOI 10.1074/jbc.M111.290569**

Zhenming Du, Jiajing Liu, Clayton D. Albracht, Alice Hsu, Wen Chen, Michelle D. Marieni, Jennie E. Williams, Julie N. Reitter, Kenneth V. Mills, and Chunyu Wang

From the Department of Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180 and the Department of Chemistry, College of the Holy Cross, Worcester, Massachusetts 01610

Protein splicing is a precise self-catalyzed process in which an intein excises itself from a precursor with the concomitant ligation of the flanking polypeptides (exteins). Protein splicing proceeds through a four-step reaction but the catalytic mechanism is not fully understood at the atomic level. We report the solution NMR structures of the hyperthermophilic *Pyrococcus abyssi* PolII intein, which has a noncanonical C-terminal glutamine instead of an asparagine. The NMR structures were determined to a backbone root mean square deviation of 0.46 Å and a heavy atom root mean square deviation of 0.93 Å. The *Pab* PolII intein has a common HINT (hedgehog intein) fold but contains an extra β-hairpin that is unique in the structures of thermo-philic inteins. The NMR structures also show that the *Pab* PolII intein has a long and disordered loop in place of an endonuclease domain. The N-terminal Cys-1 amide is hydrogen bonded to the Thr-90 hydroxyl in the conserved block-B TXXH motif and the Cys-1 thiol forms a hydrogen bond with the block F Ser-166. Mutating Thr-90 to Ala dramatically slows N-terminal cleavage, supporting its pivotal role in promoting the N-S acyl shift. Mutagenesis also showed that Thr-90 and His-93 are synergistic in catalyzing the N-S acyl shift. The block F Ser-166 plays an important role in coordinating the steps of protein splicing. NMR spin relaxation indicates that the *Pab* PolII intein is significantly more rigid than mesophilic inteins, which may contribute to the higher optimal temperature for protein splicing.

---

**Background:** The PolII intein from the hyperthermophilic *Pyrococcus abyssi* only splices at very high temperature.

**Results:** NMR structure, dynamics, and mutagenesis of *Pab* PolII intein have been characterized.

**Conclusion:** The *Pab* PolII intein has unique structural and dynamic features that may contribute to its higher temperature for optimal activity.

**Significance:** *Pab* PolII intein is an ideal candidate for protein engineering.

Protein splicing is a self-catalyzed post-translational process in which an intein excises itself from a precursor with the concomitant ligation of the flanking polypeptides (exteins). Protein splicing proceeds through a four-step reaction but the catalytic mechanism is not fully understood at the atomic level. We report the solution NMR structures of the hyperthermophilic *Pyrococcus abyssi* PolII intein, which has a noncanonical C-terminal glutamine instead of an asparagine. The NMR structures were determined to a backbone root mean square deviation of 0.46 Å and a heavy atom root mean square deviation of 0.93 Å. The *Pab* PolII intein has a common HINT (hedgehog intein) fold but contains an extra β-hairpin that is unique in the structures of thermo-philic inteins. The NMR structures also show that the *Pab* PolII intein has a long and disordered loop in place of an endonuclease domain. The N-terminal Cys-1 amide is hydrogen bonded to the Thr-90 hydroxyl in the conserved block-B TXXH motif and the Cys-1 thiol forms a hydrogen bond with the block F Ser-166. Mutating Thr-90 to Ala dramatically slows N-terminal cleavage, supporting its pivotal role in promoting the N-S acyl shift. Mutagenesis also showed that Thr-90 and His-93 are synergistic in catalyzing the N-S acyl shift. The block F Ser-166 plays an important role in coordinating the steps of protein splicing. NMR spin relaxation indicates that the *Pab* PolII intein is significantly more rigid than mesophilic inteins, which may contribute to the higher optimal temperature for protein splicing.

---

Protein splicing is a self-catalyzed post-translational process in which an intein excises itself from a precursor with the concomitant ligation of the flanking exteins, the N- and C-exteins (Fig. 1A) (1). More than 600 inteins have been found in all three domains of life: bacteria, archaea, and eukarya (2). Applications of protein splicing include protein engineering, labeling, purification, and control of protein function (3–7). Inteins can also serve as a novel drug target in bacteria that rely on protein splicing for their survival, such as *Mycobacterium tuberculosis* (8).

Protein splicing is a strictly intramolecular reaction, requiring no cofactors or energy input. The precursor protein acts as both the enzyme and substrate of the reaction. Therefore, the precursor for protein splicing is the equivalent of a traditional enzyme-substrate complex and can provide key structural information for understanding the mechanism of protein splicing. However, because of the spontaneous nature of protein splicing, native precursors are unstable. Consequently, structural studies of the protein splicing mechanism have relied on spliced inteins or intein precursors with mutations at active site residues. The *Pyrococcus abyssi* DNA polymerase II intein, abbreviated as the *Pab* PolII intein, is found in a hyperthermophilic organism that lives near deep sea thermal vents, with an optimal growth temperature of 96 °C (9). This suggests that one could isolate a stable precursor with native intein sequence for structural studies.

There are four steps of protein splicing in canonical inteins (Fig. 1A): step 1, N-X acyl shift (X = S or O); step 2, transesterification and the formation of a branched intermediate; step 3, asparagine cyclization coupled with C-terminal cleavage; and step 4, X-N acyl shift and succinimide hydrolysis (10, 11). These steps are catalyzed by active site residues in conserved blocks in intein sequences. Block A contains a conserved cysteine or serine at position 1 that serves as a nucleophile for the N-acyl shift, the first step of protein splicing. Block B has the TXXH motif important for the N-S acyl shift (12–14). Block F includes a conserved aspartate (Fig. 1A, step 3) that has been proposed to play a pivotal role in coordinating splicing and a conserved histidine that modulates C-terminal cleavage (15–20). In block G,
there is the penultimate histidine (relative to the intein C terminus) at position 6 and a C-terminal asparagine at position 7, both critical for C-terminal cleavage. The first residue of the C-extein, usually a cysteine or serine and termed C/H₁₁₀₀₁ or S/H₁₁₀₀₁, respectively, serves as the nucleophile for transesterification. Variations of the four-step scheme have been discovered among a number of inteins (21–23). The C-terminal residue of inteins is typically an asparagine, whose cyclization leads to the cleavage of the intein from the C-extein. The Pab PolII intein (23) is only the second intein demonstrated to splice with a C-terminal glutamine, after the Chiloidescent virus (CIV) RNR intein (21, 22). In contrast to the CIV RNR intein, which has reduced splicing activity upon the mutation of C-terminal glutamine to asparagine, the same mutation enhances protein splicing by 3-fold in the Pab PolII intein (23).

EXPERIMENTAL PROCEDURES

**Protein Overexpression, Purification, and NMR Sample Preparation**—The Pab PolII intein gene cloned into pETM-44 vector ppC1Q185 expresses a fusion protein with an N-terminal (His)₆ tag and maltose-binding protein (MBP). There is a single proline between the (His)₆ tag and MBP and a linker sequence TPGSLEV/LKQGPM between MBP and the intein. Isotopically labeled ([U-¹⁵Ν], [U-¹³C; U-¹⁵Ν] and [≈70%-²Η; U-¹³C; and U-¹⁵Ν]) proteins were obtained by transforming the plasmid into Escherichia coli strain BL21(DE3) and overexpressing the fusion protein in M9 medium. The M9 cultures were incubated at 37 °C until A₆₀₀ reached 0.3–0.4 and were induced with 1 mM isopropyl 1-thio-D-galactopyranoside at 20 °C for an additional 16 h. Cell lysate was purified by nickel-nickel-nitrilotriacetic acid affinity chromatography to obtain the fusion protein. The Pab PolII intein was cleaved from the fusion protein with 100 mM dithiothreitol (DTT) at 60 °C for 6 h. Affinity chromatography was utilized again to trap the (His)₆-tagged MBP and uncleaved fusion protein, whereas the Pab PolII intein was not retained by the nickel column. Flow-through fractions were pooled and exchanged into buffer containing 20 mM sodium phosphate, 0.5 mM EDTA, and 0.05 mM sodium azide in 90% H₂O, 10% D₂O or 99.9% D₂O at pH 6.5. The final concentrations of the NMR samples were between 0.5 and 2.0 mM.

**NMR Spectroscopy**—Both NMR structural and dynamics studies were carried out at 47 °C, where Pab PolII intein can mediate protein splicing efficiently (23). All spectra were acquired on a Bruker Advance II 800 MHz or Bruker Advance II 600 MHz (¹H) spectrometer, each equipped with a cryogenic probe. Spectra were processed with nmrPipe software (24) and analyzed using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, CA). Resonance assignment was carried out using the following experiments: two-dimensional [¹⁵Ν,H]-heteronuclear single quantum coherence, three-dimensional HNCACO, three-dimensional HNCO, three-dimensional HN(CO)CACB, three-dimensional [¹⁵Ν-TOCSY (~τₑ = 55 ms), three-dimensional (H)CC(CO)NH-TOCSY, three-dimen-
Solution NMR Structure of P. abyssi DNA PolII Intein

Sional H(CC)(CO)NH-TOCSY, three-dimensional HC(C)H-TOCSY ($\tau_m = 15$ ms), three-dimensional $^{15}$N-NOESY ($\tau_m = 100$ ms), and three-dimensional $^{13}$C-NOESY ($\tau_m = 105$ ms). The $^1$H chemical shifts were referenced relative to 4,4-dimethyl-4-silapentane-1-sulfonic acid and the $^{15}$N and $^{13}$C chemical shifts were referenced indirectly using frequency ratios between $^{15}$N, $^{13}$C, and $^1$H ($^{15}$N/$^1$H = 0.101329118, $^{13}$C/$^1$H = 0.251449530) (25). The chemical shifts have been deposited in the BioMagResBank under accession number 17418. For residual dipolar coupling measurements with IPAP (26), the Pab PolII intein was aligned in 7.5% polyacrylamide gels with a stretch ratio ($d_C/d_S$) of 1.29 using the apparatus described by Chou et al. (27).

$^{15}$N Relaxation Rates and Analysis—All relaxation experiments were carried out at 47 °C on a Bruker Advance II 600 MHz spectrometer equipped with a triple-resonance cryogenic probe. Relaxation properties were characterized by $^{15}$N $R_1$, $R_2$, and heteronuclear steady-state NOE experiments. These relaxation parameters are sensitive to motions occurring at the time scale faster than protein tumbling, on the order of picosecond to nanosecond. $R_1$, $R_2$, and NOE experiments were performed using the pulse sequence described by Farrow et al. (28).

NMR spectra were acquired with 2048 ($t_2$) × 256 ($t_1$) complex data points, spectral widths of 7560 Hz in $^1$H and 2798 Hz in $^{15}$N, and 16 scans. The recycle delay was 3.0 s, $R_1$ relaxation times of 10, 100, 200, 300, 400($\times$2), 500, 600, 700, 800, and 900 ms were used. $R_2$ relaxation times of 2, 16($\times$2), 30, 44, 58, 72, 86($\times$2), and 100 ms were used. $^1$H-$^{15}$N steady-state heteronuclear NOEs were obtained by interleaving the proton saturation experiment and no proton saturation experiment at each $t_1$ point. The recycle delay was 7.5 s and proton saturation was achieved by applying a 120 degree proton pulse at 5 ms delay. For each $R_1$ and $R_2$ experiment, the spectrum with the shortest relaxation time (highest intensities) was peak picked with Sparky. Fitting for relaxation rates and error estimates were accomplished using the program CURVEFIT (A. G. Palmer, Columbia University, New York). The heteronuclear NOE values were obtained from the ratio of the peak heights for $^1$H-saturated and unsaturated spectra and carried out in triplicate. The uncertainties in the NOEs were set to twice the standard deviation from three trials (29). For the Pab PolII intein, we found 104 residues with well resolved peaks in the $^{15}$N, $^1$H-heteronuclear single quantum coherence spectrum to warrant quantitative relaxation analysis on a per-residue basis. $^{15}$N NMR relaxation data were analyzed using a $r_{NOE}$ of 1.02 Å as the mean amide nitrogen-hydrogen bond length, and $\Delta\sigma = \sigma_{e} - s_{1}$ is the chemical shift anisotropy of $-172$ ppm for the backbone $^{15}$N nucleus. The amplitudes and time scales of the internal motions of the protein were determined from the relaxation data according to the model-free formalism pioneered by Lipari and Szabo (30, 31) and extended by Clore et al. (32), using the program Modelfree (version 4.15, A. G. Palmer, Columbia University) in combination with Fast Modelfree (33). The generalized order parameters ($S^2$) obtained by Modelfree described the amplitude of the internal motion for individual amide bonds at the picosecond to nanosecond time scale.

NMR Structure Determination—Peak lists were generated from three-dimensional $^{15}$N-NOESY in 90% H$_2$O, and two three-dimensional $^{13}$C-NOESY spectra in 100% D$_2$O and 90% H$_2$O, respectively, all recorded at 800 MHz $^1$H frequency. The peak lists and the chemical shifts from the resonance assignment (34) were used as input for CYANA3.0 (35). Dihedral angle restraints were derived from TALOS+ (36, 37) using chemical shifts of $^{15}$N, $^{13}$C$^\alpha$, $^{13}$C$^\beta$, $^\beta$H, and $^\alpha$H. The final set of unambiguous NOE assignments contained 3555 meaningful distance restraints, corresponding to $\sim$19 restraints/residue on average. The structure from CYANA with the lowest target function values obtained from cycle 7 were subject to refinement with residual dipolar coupling in explicit water in Xplor-NIH (38). The quality of the final structures was assessed with PSVS (39). The atomic coordinates of the bundle of 20 conformers (accession number 2LCJ) have been deposited in the Brookhaven Protein Data Bank. Model 1 represents the conformer that is closest to the mean coordinates.

Precursor Purification, Mutagenesis, and Splicing Assays—Plasmid pMIH expresses the protein MIH, previously described in Ref. 23, an in-frame fusion of E. coli MBP, the seven C-terminal residues of the native Pab PolII N-extein, the 185 residues of the intein, the five N-terminal residues of the native C-extein, and a His tag. To study the effect of mutations on the first step of splicing, we introduced mutations of C1A and Q185A to prevent steps two and three of splicing, respectively. To study the third step of splicing, we replaced the N-extein (M) with a short seven-residue polypeptide (N) and made a mutation of C1A. We overexpressed the proteins in E. coli BL21(DE3), which were purified using HisLink Protein Purification Resin (Promega, Madison, WI). The purified proteins were exchanged into buffer A (100 mM phosphate buffer, pH 7.0, with 500 mM NaCl) using 3 kDa MWCO centrifugal filters (Millipore, Billerica, MA). For the N-terminal cleavage assay, the proteins were incubated at 60 °C at the times noted in Fig. 5, B and C, in a 16- or 20-μl reaction mixture of buffer A with 2.0 μg of purified protein and supplemented with 5.0 mM EDTA, 2.0 mM tris(2-carboxyethyl)phosphine, and 100 mM DTT. For splicing or C-terminal cleavage assays, the same reaction conditions were employed, but DTT was omitted. Reactions were stopped by the addition of SDS Blue Loading Buffer (New England Biolabs, Ipswich, MA) and analyzed by SDS-PAGE using 4–20% gradient Tris glycine gels (Lonza, Rockland, ME). For Western blot analysis, gels were blotted onto PVDF. The membranes were blocked using 1% BSA in buffer W (PBS and 0.1% Tween 20) and incubated with a 1:5000 dilution of HisDetector Nickel-AP in Detector Block Solution (KPL). The blot was developed with Western Blue stabilized substrate (Promega). Conversion of MIH (precursor, 66.7 kDa) to M (43.7 kDa) and IH (23.0 kDa) indicates cleavage by DTT of the thioester linkage between the N-extein and intein. Conversion of MIH to MH (45.2 kDa) and I (21.5 kDa) indicates protein splicing, and conversion of NMIH (23.0 kDa) to NI (22.4 kDa) or MIH (66.7 kDa) to M (65.2 kDa) indicates C-terminal cleavage.

RESULTS AND DISCUSSION

NMR Structure of the Pab PolII Intein—The solution NMR structure of the Pab PolII intein is based on distance constraints, H-bond constraints, and local and long-range angular constraints, derived from NOESY, hydrogen deuterium...
exchange, chemical shift analysis using TALOS+, and residual dipolar coupling measurements, respectively (Table 1). On average, 19.5 constraints were obtained for each residue. In contrast to most inteins, the Pab PolII intein does not contain an endonuclease domain, but rather an extended loop of 26 residues (121–146) in the equivalent position in sequence (Fig. 3, highlighted in yellow). This Pab PolII-specific loop has few observed long-range NOEs and shows 15N relaxation rates characteristic of disorder and flexibility (see below). In the ensemble of 20 conformers from the Xplor-NIH calculation (38), the backbone and heavy atom root mean square deviations are 0.46 ± 0.10, 0.93 ± 0.12 Å. These values exclude the extended loop and are for residues 1–120 and 147–185 (Table 1; Fig. 3). All the residues fall into the allowed regions in the Ramachandran plot (40), with 82% in the most favored region (Table 1). Numerous structural quality factors show that the Pab PolII intein structure has better quality than the average NMR structures in PDB (Table 1) (39).

The Pab PolII intein structure is composed primarily of β-strands, which are arranged in a compact HINT (Hedgehog/Intein) fold (Fig. 2) (41). The three-dimensional structure contains 16 regular secondary structures, i.e. 14 β-strands, one α-helix, and one turn of 3_10-helix (Fig. 2B). In sequential order, the intein fold begins with a short β-strand at the N terminus, which is followed by two β-strands at residues 7–12 and 15–20. There is an amphipathic α-helix at residues 21–27, which is exposed to solvent on one side and packed against hydrophobic side chains formed by strands β2 and β3 on the other side (Leu-21, Leu-24, Tyr-25, and Leu-27). A tight turn leads to the β-hairpin of β4(31–34) and β5(37–40). This is followed by a twisted anti-parallel β-hairpin of strands β6(47–52) and extended β strand β7(59–81). β7a(59–71) forms an antiparallel β-hairpin with β5(37–40), β6(47–52), and β12b(163–167), whereas β7b(76–81) forms an antiparallel β-hairpin with β8(86–99) and β12a(149–157). The strand β9(94–99) pairs with β10(102–107) to from a hairpin, which is then connected

### Table 1

The experimental NMR data for the structure calculation and the structural statistics of the 20 energy-minimized conformers of Pab PolII intein

| (a) NMR constraints | Value |
|---------------------|-------|
| Distance constraints| 3231  |
| intra-residue | 774   |
| sequential | 843   |
| medium range | 345   |
| long range | 1269  |
| Dihedral-angle constraints | 223   |
| ψ | 110   |
| Hydrogen bonds | 51    |
| Residual dipolar couplings | 90    |
| Total number of restricting constraints | 3595  |

| (b) RMSD from ideal geometry |
|-----------------------------|
| Bond lengths, Å | 0.016 |
| Bond angle, ° | 0.9   |

| (c) RMSD to mean coordinates, Å |
|-------------------------------|
| All backbone atoms excluding disordered loop residues 121–146 | 0.46 ± 0.10 |
| All heavy atoms excluding disordered loop residues 121–146 | 0.93 ± 0.12 |

| (d) Ramachandran plot statistics^a |
|-----------------------------------|
| Residues in most favored regions  | 82.0% |
| Residues in additionally allowed regions | 15.4% |
| Residues in generally allowed regions  | 2.6% |
| Residues in disallowed regions | 0.0% |

| (e) Structure quality factors |
|------------------------------|
| Verify3D | Z-Score of Pab PolII intein structure^c | Z-Score of pdb_nmr^c |
| 0.43 | -0.48 | -2.15 |
| Procheck (phi-psi) | -0.67 | -0.32 | -2.55 |
| Procheck (all dihedrals) | -0.65 | -3.73 | -5.08 |
| MolProbity:Cashscore | 54.5 | -7.82 | -10.74 |

^a Residues for Ramachandran statistics are 5–90, 93–120, 147–158, 161–170, and 174–183.

^b With respect to mean and SD for a set of 252 x-ray structures <500 residues, of resolution < 1.80 Å, R-factor < 0.25 and R-free < 0.28; a positive value indicates a “better” score.

^c 183 NMR structures determined by nonstructural genomics consortia groups.

**FIGURE 2.** A, ensemble of 20 conformers representing the Pab PolII intein structure. B, ribbon presentation of the Pab PolII intein.
to a short 3_{10} helix. B11(116–119) is connected to the long Pab PolII-specific loop that passes through the center of the disk-shaped intein structure with the active site residue Ser-166 located in 12b(163–167). Additional 13(175–177) and 14(181–183) strands form an antiparallel -hairpin and bring the C-terminal Gln to the active site. Strands 7 and 12 are continuous in some inteins but in the Pab PolII intein appear to be broken by short loops (Fig. 3), and are therefore named as 7a(59–71), 7b(76–81), 12a(149–157), and 12b(163–167).

A Unique -Hairpin in Thermophilic Inteins (HTH)—A structure-based sequence alignment of Pab PolII with homologous proteins (inteins and hedgehog processing domains) using the DALI server. PDB accession codes are listed after each protein. The residue numbers and regular secondary structure elements are indicated above the sequences. The locations of the conserved intein sequence blocks A, B, F, and G are indicated above the sequence with corresponding residues underlined. Key residues for protein splicing are marked in red. Hyphens (-) indicate gaps in the alignment. The proposed hyperthermophilic hairpin is shadowed in green and an extended loop that replaces the endonuclease domain is shadowed in yellow. Mutations are shadowed in gray.
The three-dimensional structures of the Pab PolII intein superimpose nicely with the Tko Pol-2, Mja KlbA, Mxe GyrA, and Mtu RecA inteins, as shown in Fig. 4, A–D, respectively. In addition to the HINT domain, the Tko Pol-2 intein contains the endonuclease domain, domains III and IV (Fig. 4A), whereas the Pfu RIR1–1 intein contains the endonuclease domain and the extra Stirrup domain (not shown). In inteins, the endonuclease domain is generally inserted in the HINT domain between /H925211 and /H925212. In the Pab PolII intein, however, there is a unique and extended loop between /H925211 and /H925212 instead of an endonuclease domain. The extended loop does not affect the HINT fold of the Pab PolII intein, with low Cα root mean square deviation values (1.6 Å for 185 aligned residues of Pab PolII versus Tko Pol-2 (2CW7) inteins; 1.7 Å for 185 aligned residues of Pab PolII versus Pfu RIR1–1 (1DQ3) inteins).

Structure-based sequence alignment also shows that the four thermophilic archaean inteins, Pab PolII, Tko Pol-2, Pfu RIR1–1, and Mja KlbA have insertions of ~18 residues between positions 29 and 46, relative to other inteins. These insertions form a β-hairpin, composed of the β4 and β5 strands that are connected by a short loop. This β-hairpin is colored blue in Fig. 4. Such a β-hairpin is only present in the structures of thermophilic inteins among inteins with known three-dimensional structures, such as the Tko Pol-2 intein (Fig. 4A) (43), Mja KlbA intein (Fig. 4B), and Pfu RIR1–1 intein (44) (not shown). This β-hairpin is missing in the structures of mesophilic inteins, such as the Mtu RecA intein (18)(Fig. 4D). We therefore propose to name it the HTH. We speculate that the HTH could enhance the stability of thermophilic inteins by extending the β-sheet by two strands and may contribute to the higher optimal splicing temperature of thermophilic inteins.

Interestingly, the mesophilic Mxe GyrA intein (12) contains a short segment that corresponds to /H92525 in Pab PolII but is missing the equivalent of /H92524 (Fig. 4C).

Comparison of the Pab PolII intein with three other archaean inteins shows that they all contain an α-helix in positions corresponding to residues 21–27 in the Pab PolII intein. This helix is longer in the Pab PolII intein than in the HINT domain of other inteins, such as Ssp DnaB (16), Mxe GyrA (12), Sce VMA (13), Ssp DnaE (41), and Mtu RecA (18).

Structure-Function Relationship of Active Site Residues—The active site is composed of Cys-1, block B TXXH (residue 90 to
93), block F Ser-166, and the C-terminal Gln-185 (Fig. 5A). Both side chains of Thr-90 and His-93 are close to Cys-1. The Thr-90 hydroxyl is well positioned to serve as a hydrogen bond donor to the Cys-1 amide nitrogen (Fig. 5A). Ser-166 in the Pab PolII intein is at the equivalent position as the conserved block F aspartate, which plays a coordinating role in protein splicing.
The Ser-166 side chain forms a hydrogen bond with the Cys-1 thiol, with a distance of 3.8 Å between its O\(^\cdot\) and the sulfur atom of Cys-1. This hydrogen bond is similar to the one observed between the block F aspartate and Cys-1 in the Mtu RecA intein (20). In the NMR structure of the Pab PolII intein, the C-terminal Gln-185 is not well defined. Gln-185 was not observed in the \(^{15}\text{N},^{1}\text{H}\)-heteronuclear single quantum coherence spectrum, resulting in a few structural constraints. It is likely that Gln-185 experiences microsecond to millisecond time scale motion, which results in the broadening and disappearance of its NMR signal.

We have constructed a series of mutants of block B TXXH and block F Ser to test their role in the catalytic mechanism of the Pab PolII intein. To detect splicing or N-terminal cleavage, we used an MIH construct, with M, I, and H representing the N-extein (MBP plus seven native extein residues), intein (Pab PolII intein), and C-extein (five native extein residues plus a His tag), respectively. To study the isolated C-terminal cleavage, we used an NIH construct where the MBP was deleted from the expression context and the N-extein consists of just seven residues.

To study the role of the TXXH motif (residues 90–93) in the N-S acyl shift, we constructed a Q185A/C+1A double mutant so that the precursor can only undergo N-terminal cleavage to yield M and IH. We incubated the protein in a neutral buffer at 60 °C with a saturating concentration of DTT such that N-terminal cleavage approximates the rate and extent of thioester formation.

The effects of T89A, T90A, D92A, H93A, and a T90A/H93A double mutant on N-terminal cleavage are shown in Fig. 5B. With the wild type TXXH motif (in the Q185A/C+1A double mutant background), there are visible bands of both M and IH after 1 h of incubation with DTT, with the concomitant decrease in precursor MIH intensity, indicating the occurrence of the N-S acyl shift and DTT-mediated N-terminal cleavage. The band intensity in the cleavage products continues to increase with time, whereas the intensity for precursor continues to decrease. By 16 h, there is little precursor remaining. Mutating the nonconserved Thr-89 and Asp-92 to Ala slightly increases the rate of N-terminal cleavage. Mutation of either Thr-90 or His-93 to Ala dramatically slows the rate of N-terminal cleavage; substantial precursor remains even after 16 h. The T90A/H93A double mutant abolishes N-terminal cleavage, with no visible band of either M or IH after 16 h. Thr-90 is as important as His-93 in catalyzing the N-S acyl shift in the Pab PolII intein. The biochemical role of Thr-90 is supported by the structure. The catalytic mechanism of Thr-90 may be due to a hydrogen bond between the Thr-90 hydroxyl oxygen atom and the Cys-1 amide nitrogen, which are separated by 2.9 Å. This hydrogen bond may stabilize the negatively charged oxythiazolidine intermediate. The Cys-1 carbonyl is within 2.9 Å of the Thr-90 hydroxyl and 3.2 Å of the Thr-90 amide nitrogen; these hydrogen bonds may serve to properly orient Cys-1 in the active site to facilitate the nucleophilic attack of the Cys-1 thiol. Alternatively, Thr-90 may adopt different conformations in the unspliced precursor due to the presence of exteins. Thr-90 may form a hydrogen bond with the \(-1\) carboxyl, which has been observed in the Mja KlBA (45) and Mxe GyrA inteins (12). This hydrogen bond could stabilize the negatively charged carboxyl oxygen of the oxythiazolidine anion. The imidazole \(\epsilon\)-nitrogen of His-93 is within 3.5 Å of the Cys-1 thiol in three of the 20 NMR conformers so that the Cys-1 thiol may be deprotonated by block B histidine for the initiation of the N-S acyl shift. In eight of the 20 NMR conformers, the imidazole \(\epsilon\)-nitrogen of His-93 is within 3.5 Å of the Cys-1 amide nitrogen so that the B block histidine may protonate the leaving group (Cys-1 amide) during the N-S acyl shift. These structural observations are consistent with the proposed dual catalytic role for the B block histidine in the N-S acyl shift (52).

We also studied the influence of mutations at Ser-166 on both N- and C-terminal cleavage. The equivalent residue of Ser-166 in other inteins, such as Asp-422 in the Mtu RecA intein, plays a coordination role in protein splicing (18–20, 45). In Fig. 5C, we observe that the rate of isolated N-terminal cleavage is reduced with mutation of Ser-166 to Gly, Asp, or Ala, but remains unchanged with mutation to Thr. This suggests a role in promoting the first step of splicing for the hydrogen bond between the Ser-166 hydroxyl and Cys-1 thiol, which are separated by 3.8 Å (Fig. 5A). Ser-166 also plays a role in regulating C-terminal cleavage. In Fig. 5D, we note that mutation of Ser-166 to Gly greatly accelerates the rate of C-terminal cleavage due to Asn cyclization, and mutation of Ser-166 to Asp slows the rate dramatically. Similar results are observed with the native C-terminal Gln in Fig. 5E. These results are analogous to the Asp-422 to Gly mutation ("the C-terminal cleavage mutation") in the Mtu RecA intein, which promotes C-terminal cleavage at the expense of splicing (17).

The effects on protein splicing of Ser-166 mutants within the context of a Gln-185 to Asn mutation are shown in Fig. 5F. The identities of the major bands were also confirmed by MALDI-TOF mass spectrometry and Western blot. Mutating Ser-166 to the more frequently observed Asp results in some splicing, but mostly N-terminal cleavage uncoupled from splicing. Mutating Ser-166 to Gly abolishes protein splicing and yields exclusively C-terminal cleavage product, as also observed in the D422G cleavage mutant of the Mtu RecA intein (17). Mutation of Ser-166 to Ala or Thr promotes mostly uncoupled N- and C-terminal cleavage, along with some splicing. The block F Ser is therefore vital to coupling the reactions at the N- and C-terminal splice junctions of the intein.

\(^{15}\text{N}\) Spin Relaxation—Both conformational changes and protein dynamics are important in intein catalysis (15, 41). We measured the \(^{15}\text{N} R_{1}, R_{2}, \) and heteronuclear NOE and performed model-free analysis using Fast Modelfree (33) to obtain the order parameters.

As shown in Fig. 6, most residues exhibit high order parameters, indicating that the majority of the protein, including the N and C termini, is well ordered. The residues of the active site, such as Thr-90, His-93, and Ser-166 show no observable differences in relaxation from the rest of the protein. Residues in the extended loop region (121–146) show reduced heteronuclear NOE values (<0.6), lower \(R_{2}\) values, and slightly higher \(R_{1}\) values than the rest of the protein, indicating large-amplitude
picosecond to nanosecond motions, consistent with the lack of long-range NOE cross-peaks and increased disorder in the NMR structure (Fig. 2A). There is also increased mobility in the loop connecting strands β12a and β12b, with residues 159 and 161 showing decreased order parameters. The average order parameter in the HTH (0.92 ± 0.02) is similar to that of the rest
of the protein (0.93 ± 0.02, excluding residues from the extended loop 121–146). One plausible explanation is that the HTH stabilizes the entire Pab PolII intein, instead of the HTH alone.

The Pab PolII and Mtu RecA inteins (53) have a very similar pattern of order parameters for structurally conserved residues, with average order parameters of 0.93 (excluding residues from the extended loop 121–146) and 0.89, respectively. Both inteins are more rigid than typical proteins where the average order parameter is ~0.85 for structured regions (α-helix and β-strands) (54). The Pab PolII intein is even more rigid than the Mtu RecA intein. Structure-based comparison of order parameters, which compares paired residues in the Pab PolII intein and the Mtu RecA intein using structure-based alignment (Fig. 3), shows that the Pab PolII intein has ~4% higher order parameters on average. In addition, relaxation experiments were carried out at 47 °C for the Pab PolII intein, whereas they were measured at a lower temperature of 25 °C for the Mtu RecA intein. As lower temperature increases rigidity, the Pab PolII intein is expected to be much more rigid than the Mtu RecA intein at 25 °C. In contrast, similar order parameters were measured at a lower temperature of 25 °C for the PolII intein, whereas they were 30–40% lower than the rest of the protein (56). This unusual rigidity was also observed for terminal residues of the intein in Mja KlBA (45) and Npu DnaE (57), but not for terminal residues in the N- or C-exteins.

The extended loop in the Pab PolII intein, characterized by disorder and mobility, could provide a site for protein engineering, such as creating an artificially split intein or the control of protein splicing and functions with small molecules. Artificially split inteins from the Pab PolII intein might require an elevated temperature for optimal splicing, which could increase the temperature range for the applications of intein trans-splicing. Inteins have also been engineered to respond to small molecules to control protein function (58). A ligand binding domain could be inserted into the disordered loop in the Pab PolII intein such that it can be engineered to respond to pH (59), light (60), temperature (61), or small molecules (62) through directed evolution.

Acknowledgments—We thank Lauren Duffee, Colleen Donahue, and Deirdre Dorval for experimental help.

REFERENCES
1. Paulus, H. (2000) Annu. Rev. Biochem. 69, 447–496
2. Perler, F. B. (2002) Nucleic Acids Res. 30, 383–384
3. Kwon, Y., Coleman, M. A., and Camarero, J. A. (2006) Angew. Chem. Int. Ed. 45, 1726–1729
4. Seyedsayamdost, M. R., Yee, C. S., Reece, S. Y., Nocera, D. G., and Stubbe, J. (2006) J. Am. Chem. Soc. 128, 1562–1568
5. Evans, T. C., Jr., Benner, J., and Xu, M. Q. (1999) J. Biol. Chem. 274, 18359–18363
6. Züger, S., and Iwai, H. (2005) Nat. Biotechnol. 23, 736–740
7. Romanelli, A., Shekhtman, A., Cowburn, D., and Muir, T. W. (2004) Proc. Natl. Acad. Sci. U.S.A. 101, 6397–6402
8. Paulus, H. (2007) Drug Future 32, 973–984
9. Eruaso, G., Reysenbach, A. L., Godfrey, A., Meunier, J. R., Crump, B., Partensky, F., Baross, J. A., Martineinson, V., Barbier, G., Pace, N. R., and Prieur, D. (1993) Arch. Microbiol. 160, 338–349
10. Paulus, H. (1998) Chem. Soc. Rev. 27, 375–386
11. Saleh, L., and Perler, F. B. (2006) Chem. Rev. 106, 183–193
12. Klabunde, T., Sharma, S., Tenleti, A., Jacobs, W. R., Jr., and Sicchettini, J. C. (1998) Nat. Struct. Biol. 5, 31–36
13. Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y., and Satow, Y. (2002) J. Mol. Biol. 316, 919–929
14. Mizutani, R., Anraku, Y., and Satow, Y. (2004) J. Synchrotron Radiat. 11, 109–112
15. Frutos, S., Goger, M., Giovani, B., Cowburn, D., and Muir, T. W. (2010) Nat. Chem. Biol. 6, 527–533
16. Ding, Y., Xu, M. Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2003) J. Biol. Chem. 278, 39133–39142
17. Wood, D. W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999) Nat Biotechnol. 17, 889–892
18. Van Roey, P., Pereira, B., Li, Z., Hiraga, K., Belfort, M., and Derbyshire, V. (2007) J. Mol. Biol. 367, 162–173
19. Pereira, B., Shemella, P. T., Amitai, G., Belfort, G., Nayak, S. K., and Belfort, M. (2011) J. Mol. Biol. 406, 430–442
20. Du, Z., Zheng, Y., Patterson, M. N., Liu, Y., and Wang, C. (2011) J. Am. Chem. Soc. 133, 10275–10282
21. Amitai, G., Dassa, B., and Pietrokovski, S. (2004) J. Biol. Chem. 279, 3121–3131
22. Pietrokovski, S. (1998) Curr. Biol. 8, R634–635
23. Mills, K. V., Manning, J. S., Garcia, A. M., and Wuerdeman, L. A. (2004) J. Biol. Chem. 279, 20685–20691
24. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) J. Biomol. NMR 6, 277–293
25. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) J. Biol. Chem. 270, 135–140
26. Ottiger, M., Delaglio, F., and Bax, A. (1998) J. Magn. Reson. 131, 373–378
27. Chou, J. J., Gaemers, S., Hawder, B., Louis, J. M., and Bax, A. (2001) J. Biomol. NMR 21, 377–382
28. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascall, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Biochemistry 33, 5984–6003
29. Savard, P. Y., and Gagné, S. M. (2006) Biochemistry 45, 11414–11424
30. Lipari, G., and Szabo, A. (1981) J. Am. Chem. Soc. 103, 4549–4570
31. Lipari, G., and Szabo, A. (1982) J. Am. Chem. Soc. 104, 4546–4559
32. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) J. Am. Chem. Soc. 112, 4989–4991
33. Cole, R., and Loria, J. P. (2003) J. Biomol. NMR 26, 203–213
34. Liu, J., Du, Z., Albracht, C. D., Naidu, R. O., Mills, K. V., and Wang, C. (2011) Biomol. NMR Assign.
35. Güntert, P. (2004) Methods Mol. Biol. 278, 353–378
36. Cornilescu, G., Delaglio, F., and Bax, A. (1999) J. Biomol. NMR 13, 289–302
37. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) J. Biomol. NMR 44, 213–223
38. Schwiers, C. D., Kuszewski, J. J., and Clore, G. M. (2006) Prog. Nucl. Mag. Res. Sp. 48, 47–62
39. Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007) Proteins 66, 778–795
40. Ramachandran, G. N., Ramakrishnan, C., and Sasiakevich, V. (1963) J. Mol. Biol. 7, 95–99
41. Sun, P., Ye, S., Ferrandon, S., Evans, T. C., Xu, M. Q., and Rao, Z. (2005) J. Mol. Biol. 353, 1093–1105
42. Holm, L., and Sander, C. (1993) J. Mol. Biol. 233, 123–138
43. Matsumura, H., Takahashi, H., Inoue, T., Yamamoto, T., Hashimoto, H., Nishioka, M., Fujisawa, S., Takagi, M., Imanaka, T., and Kai, Y. (2006) Proteins 63, 711–715
Solution NMR Structure of P. abyssi DNA PolII Intein

44. Ichiyang, K., Ishino, Y., Ariyoshi, M., Komori, K., and Morikawa, K. (2000) *J. Mol. Biol.* **300**, 889–901
45. Johnson, M. A., Southworth, M. W., Herrmann, T., Brace, L., Perler, F. B., and Wüthrich, K. (2007) *Protein Sci.* **16**, 1316–1328
46. Werner, E., Wende, W., Pingoud, A., and Heinemann, U. (2002) *Nucleic Acids Res.* **30**, 3962–3971
47. Moure, C. M., Gimble, F. S., and Quiocho, F. A. (2002) *Nat. Struct. Biol.* **9**, 764–770
48. Hu, D., Crist, M., Duan, X., Quiocho, F. A., and Gimble, F. S. (2000) *J. Biol. Chem.* **275**, 2705–2712
49. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) *Cell* **89**, 555–564
50. Poland, B. W., Xu, M. Q., and Quiocho, F. A. (2000) *J. Biol. Chem.* **275**, 16408–16413
51. Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) *Cell* **91**, 85–97
52. Du, Z., Shemella, P. T., Liu, Y., McCallum, S. A., Pereira, B., Nayak, S. K., Belfort, G., Belfort, M., and Wang, C. (2009) *J. Am. Chem. Soc.* **131**, 11581–11589
53. Du, Z., Liu, Y., Ban, D., Lopez, M. M., Belfort, M., and Wang, C. (2010) *J. Mol. Biol.* **400**, 755–767
54. Palmer, A. G., 3rd (2001) *Annu. Rev. Biophys. Biomol. Sci.* **30**, 129–155
55. Butterwick, J. A., Patrick Loria, J., Astrof, N. S., Kroenke, C. D., Cole, R., Rance, M., and Palmer, A. G., 3rd (2004) *J. Mol. Biol.* **339**, 855–871
56. Goodman, J. L., Pagel, M. D., and Stone, M. J. (2000) *J. Mol. Biol.* **295**, 963–978
57. Oeemig, J. S., Aranko, A. S., Djupsjöbacka, J., Heinämäki, K., and Iwai, H. (2009) *FEBS Lett.* **583**, 1451–1456
58. Buskirk, A. R., Ong, Y. C., Gartner, Z. J., and Liu, D. R. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10505–10510
59. Wood, D. W., Derbyshire, V., Wu, W., Chartrain, M., Belfort, M., and Belfort, G. (2000) *Biotechnol. Prog.* **16**, 1055–1063
60. Tyszkiewicz, A. B., and Muir, T. W. (2008) *Nat. Methods* **5**, 303–305
61. Zeidler, M. P., Tan, C., Bellaiche, Y., Cherry, S., Häder, S., Gayko, U., and Perrimon, N. (2004) *Nat. Biotechnol.* **22**, 871–876
62. Mootz, H. D., Blum, E. S., Tyszkiewicz, A. B., and Muir, T. W. (2003) *J. Am. Chem. Soc.* **125**, 10561–10569