Biochemical and Biophysical Research Communications 403 (2010) 457–461

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Mechanism of protein splicing of the *Pyrococcus abyssi* lon protease intein

Kevin M. O'Brien, Ann K. Schufreider, Melissa A. McGill, Kathryn M. O'Brien, Julie N. Reitter, Kenneth V. Mills *

Department of Chemistry, College of the Holy Cross, Worcester, MA 01610, USA

ARTICLE INFO

Article history:
Received 5 November 2010
Available online 19 November 2010

Keywords:
Protein splicing
Intein
Extein
Pyrococcus abyssi
Lon protease
Homing endonuclease

ABSTRACT

Protein splicing is a post-translational process by which an intervening polypeptide, the intein, excises itself from the flanking polypeptides, the exteins, coupled to ligation of the exteins. The lon protease of *Pyrococcus abyssi* (*Pab*) is interrupted by an intein. When over-expressed as a fusion protein in *Escherichia coli*, the *Pab* lon protease intein can promote efficient protein splicing. Mutations that block individual steps of splicing generally do not lead to unproductive side reactions, suggesting that the intein tightly coordinates the splicing process. The intein can splice, although it has Lys in place of the highly conserved penultimate His, and mutants of the intein in the C-terminal region lead to the accumulation of stable branched-ester intermediate.

1. Introduction

Protein splicing is a post-translational process by which an intervening polypeptide, called an intein, facilitates its own excision from the flanking polypeptides, or exteins, and the ligation of the exteins.

The canonical mechanism of protein splicing involves four steps [1,2]. The first step is nucleophilic attack of the N-terminal intein residue, usually Ser or Cys, on the peptide bond linking the N-extein and intein to create an ester (or thioester). (See Fig. 1 for a schematic of the *Pyrococcus abyssi* lon intein.) Step two is a transesterification by which the N-extein is transferred from the side chain of the first intein residue to the side chain of the first C-extein residue. In step three, the intein is cleaved from the ligated exteins via cyclization of its C-terminal Asn coupled to peptide bond cleavage. In step four, the ester bond linking the exteins is converted to the more stable peptide bond, and the intein C-terminal aminosuccinimide is converted to Asn or iso-Asn.

We chose to study the intein that interrupts the lon protease of the hyperthermophile *P. abyssi* because it lacks the highly conserved penultimate His and is flanked at its C terminus by a Gly-rich region, which may provide needed conformational flexibility to promote the third step of splicing. We show that the *Pab* lon intein can promote protein splicing when expressed as a fusion protein in *Escherichia coli*. Generally, the intein does not promote side reactions if the overall splicing process is disrupted by mutation. This is unusual, as most inteins will promote N-terminal cleavage of the ester from step one or two if Asn cyclization is prevented, or promote C-terminal cleavage via Asn cyclization uncoupled from splicing if step one and/or step two is slowed or blocked [2].

2. Materials and methods

2.1. Plasmid preparation

To amplify the intein gene via PCR, *P. abyssi* was cultured and genomic DNA isolated as described [3]. The intein was amplified by PCR using oligonucleotide primers lonU (5′-CCGGAAGGAAGAGGCCTCCATTCGT) and lonL (5′-TGCTATCTCATCGATGAAAAGGACG).

To generate a plasmid that codes for a fusion protein consisting of *E. coli* maltose binding protein (MBP), the intein and a poly-His tag, plasmid pMIHlon was created by digesting the PCR product above with StuI and ClaI and ligating the product of the digestion into the same sites in plasmid pPabPol1His, described previously [3].

To generate a plasmid that codes for the intein with an N-terminal poly-His tag and MBP and a C-terminal glutathione-*S*-transferase tag (GST), the DNA cassette flanked by the restriction sites for StuI and ClaI from plasmid pHMRIR1 (previously described [4]) was replaced with the StuI/ClaI cassette from plasmid pMIHlon. Mutations described were generated using appropriate oligonucleotide primers via site directed mutagenesis with PfuTurbo DNA polymerase (Agilent Technologies, Santa Clara, CA).

The DNA sequences of all plasmids were confirmed by Macrogen, Inc. (Seoul, Korea). The DNA sequence of the intein was consistent with the NCBI database (Accession No. AJ248288).

* Corresponding author. Fax: +1 508 793 3530.
E-mail address: kmills@holycross.edu (K.V. Mills).

0006-291X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2010.11.055

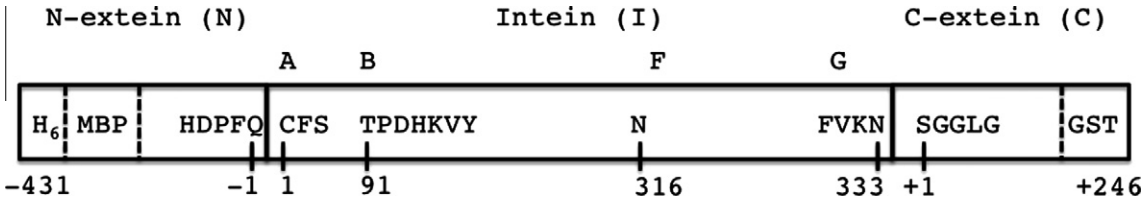

Fig. 1. Schematic illustration of the *Pab* lon intein. The fusion protein NIC consists of a poly-His tag and maltose binding protein fused to the 23 C-terminal residues of the native N-extein, the 333-residue intein, the 27 N-terminal residues of the native C-extein, and a glutathione-*S*-transferase tag. The residue numbering scheme is below the schematic and the conserved intein blocks (A, B, F, G) are above [33]. The schematic is not to scale.

2.2. Expression and protein purification

Proteins were over-expressed in *E. coli* BL21DE3 cells (EMD, Gibbstown, NJ). Cultures (50 mL) were grown at 37 °C with shaking to mid-log phase, induced with isopropyl-1-thio-β-D-galactopyranoside (final concentration of 1 mM), and grown overnight with shaking at 20 °C. The cells were harvested by centrifugation and resuspended in 2.75 mL of Buffer A (100 mM bis-Tris propane, pH 7.5, 500 mM NaCl) supplemented with 12 units/ml benzonase nuclease (EMD), 100 μM phenylmethylsulfonyl fluoride and 20 μl of Protease Inhibitor Cocktail P8849 (Sigma–Aldrich, St. Louis, MO). The cells were disrupted by BugBuster reagent (EMD). The clarified supernatant was applied to a 1 mL slurry of 50% amylose resin (New England Biolabs, Ipswich, MA) pre-equilibrated with Buffer A. The resin was washed with 10 ml of Buffer A, 10 mL of Buffer A supplemented with 0.1% Tween 20, and 10 mL of Buffer A with 1.0 M NaCl. The proteins were eluted in three 500-μl fractions of Buffer A supplemented with 20 mM maltose. Protein concentrations were determined by the Bradford method [5].

To change the pH of the elution fraction, buffer was exchanged against 100 mM MES, pH 5.5, 100 mM sodium phosphate, pH 7.0, or 100 mM HEPES, pH 8.5, each with 500 mM NaCl, using a Millipore Ultrafree-0.5 centrifugal filter device.

2.3. Protein splicing analysis by SDS–PAGE, Western blot, N-terminal sequencing and MALDI-TOF mass spectrometry

Proteins were analyzed by SDS–PAGE using pre-cast 4–20% gradient Tris–glycine gels (Lonza, Rockland, ME) stained with Coomassie blue.

For Western blot analysis, gels were blotted onto PVDF (Millipore). The membranes were blocked using 1% bovine serum albumin in Buffer W (PBS and 0.1% Tween 20) and then incubated with a 1:4000 dilution of His-detector Nickel-AP conjugate (KPL, Gaithersburg, MD) or a 1:10,000 dilution of GST·Tag monoclonal antibody (EMD). The blots were washed in Buffer W. The anti-GST blot was incubated with a 1:8000 dilution of anti-mouse IgG alkaline phosphatase conjugate (EMD) and washed again. Blots were developed with Western Blue stabilized substrate (Promega, Madison, WI).

For N-terminal sequencing, gels were blotted onto PVDF as described [4] and analyzed by the Tufts University Core Facility (Boston, MA). MALDI-TOF mass spectrometry data were collected by the Tufts University Core Facility.

3. Results

3.1. Protein expression

The lon protease of *P. abyssi* is interrupted by an intein [6]. In *E. coli*, we over-expressed an intein fusion protein that consists of an N-terminal poly-His tag followed by *E. coli* MBP, the 23 C-terminal residues of the N-extein, the 333 residues of the *Pab* lon protease intein, the 27 N-terminal residues of the C-extein, and a C-terminal GST. Although the *Pab* lon intein is interrupted by a homing endonuclease domain, the fusion protein expresses well. This is consistent with the observation that this homing endonuclease is not functional [7], as intein fusion proteins with active homing endonuclease domains can express poorly in *E. coli* [4]. A schematic of the intein fusion protein and residues examined is given in Fig. 1.

3.2. The overall splicing reaction

Unlike the *Pab* PolII intein [3], the *Pab* lon intein splices essentially to completion when over-expressed in *E. coli* (Fig. 2). Blocking the first step of splicing with a Cys1 to Ala mutation or step two of splicing with a Ser+1 to Ala mutation leads to the accumulation of precursor (Fig. 2). Substitution of Ser+1 with Thr results in precursor, whereas substitution to Cys results in efficient splicing (Fig. 2). Prevention of step three of splicing by mutation of the intein C-terminal Asn333 to Ala results in the accumulation of precursor and branched-ester intermediate.

We identified the band in lane N333A of Fig. 2 as branched-ester intermediate based on N-terminal sequencing that results in two over-lapping N-terminal sequences attributable to the N-extein and intein (except for the N-terminal Cys of the intein, which is not detectable by this method). In Fig. 2 and each subsequent experiment, SDS–PAGE bands were confirmed by Western blot. Bands were identified as BE (branched ester), NIC (precursor), NC (spliced product), or N (N-terminal product of N-terminal cleavage) using an antibody directed against the His tag in the N-extein (Fig. 2b). They also were identified as BE, NIC, NC, or IC (C-terminal product of N-terminal cleavage) using an antibody directed against GST in the C-extein (Fig. 2c).

To verify the identity of the splice junction, we analyzed the spliced product of the wild-type intein fusion protein by MALDI-TOF mass spectrometry. The *m*/*z* was 75,820, within 0.4% of the expected Mr of 75,521. We also analyzed the precursor fusion protein with a Cys1Ala mutation by MALDI-TOF MS. The *m*/*z* was 114,217, within 0.6% of the expected Mr of 113,513. We routinely detect a peak at a *m*/*z* of about 56,850, consistent with a band seen equally in each lane that we attribute to non-specific binding to the resin.

As there is a Ser at position three of the intein, we were concerned that it could serve as an alternate N-terminal nucleophile. However, an intein fusion protein with a Ser3Ala mutation splices, albeit with reduced efficiency (Fig. 4). Because the intein with a Cys1Ala mutation cannot splice, and the Ser3Ala mutant does, we believe that this Ser may help to coordinate the active site but is not involved as a nucleophile.

3.3. Further analysis of the branched ester intermediate

Mutation of Asn333 to Ala results in the accumulation of precursor and branched-ester intermediate (Figs. 2 and 4). A band of higher molecular weight also appears in lanes with branched ester,