Available online at www.sciencedirect.com
ScienceDirect
Biochemical and Biophysical Research Communications 355 (2007) 830–834
BBRC
www.elsevier.com/locate/ybbrc

Trans-splicing of an artificially split fungal mini-intein

Skander Elleuche, Stefanie Pöggeler *

Institute of Microbiology and Genetics, Department of Genetics of Eukaryotic Microorganisms, Georg-August-University of Göttingen, Germany

Received 6 February 2007
Available online 15 February 2007

Abstract

Inteins are internal protein domains found inside the coding region of different proteins. They can autocatalytically self-excise from their host protein and ligate the protein flanks, called exteins, with a peptide bond via a post-translational process called protein *cis*-splicing. In contrast, protein *trans*-splicing involves inteins split into an N- and a C-terminal domain. Both domains are synthesized as two separate components and each joined to an extein; the intein domains can reassemble and link the joined exteins into one functional protein. In this study, we introduced three split sites into the PRP8 mini-intein of *Penicillium chrysogenum* and demonstrated for the first time *trans*-splicing of a fungal PRP8 intein. Two of the sites introduced allowed splicing to occur in *trans* while the third was not functional.

Keywords: Intein; Protein splicing; *Penicillium chrysogenum*; PRP8 protein; *Trans*-splicing

An intein is a self-catalytic protein-intervening sequence that catalyzes its precise excision from a host protein and the ligation of its flanking sequences, termed N- and C-exteins, to produce a mature spliced product [1]. Protein splicing is a post-translational event that releases an internal intein sequence from a protein precursor. The mechanism of protein splicing typically consists of four steps: two acyl rearrangements at the two splicing junctions, a *trans*-esterification between the two junctions, and cyclization of a residue at the C-terminal junction. The splicing reaction depends on three highly conserved residues with the vast majority of inteins beginning with a serine or cysteine residue, and having an asparagine residue at the C-terminus and a hydroxyl- (serine) or thiol- (cysteine) containing residue at the first position of the C-terminal extein [2–4]. Many inteins are bi-functional and contain a homing endonuclease domain as a second discrete functional region [5,6]. Inteins are termed large or minimal (mini-) depending on whether they contain an endonuclease domain or not [5]. The endonuclease domain of inteins is not involved in protein splicing which is carried out only by the splicing domain. The splicing domain of large inteins appears to be split by the endonuclease domain into an N-terminal and a C-terminal splicing region that both contain conserved blocks of amino acids, i.e., blocks A and B and blocks G and F for the N- and C-terminal splicing regions, respectively [7,8]. X-ray structures of diverse inteins have revealed that inteins form a common horseshoe-like structural β-strand scaffold known as the Hint (Hedgehog, intein) motif [9–13].

Several inteins exist as two fragments and are encoded by two separately transcribed and translated genes. These so-called split inteins self-associate and catalyze protein-splicing activity in *trans*. Native split inteins have been identified in diverse cyanobacteria and the archeaon *Nanoarchaeum equitans* [14–17] or have been artificially engineered from *cis*-splicing inteins [18–20]. Naturally split inteins as well as engineered split inteins can be used for various applications in protein chemistry, e.g., protein ligation, peptide cyclization, regulation of protein activity, and analysis of protein–protein interactions [21].

In fungal organisms, only *cis*-splicing inteins have been detected. These are all located within nuclear-encoded

* Corresponding author. Fax: +49 551 3910123.
E-mail address: spoegge@gwdg.de (S. Pöggeler).

0006-291X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2007.02.035

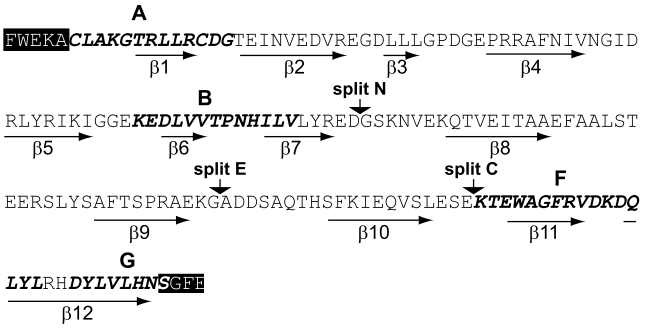

Fig. 1. Amino acid sequence of the 157-aa PRP8 mini-intein from *P. chrysogenum*. Positions of three split sites (split N, split E, and split C) are marked. Flanking extein regions are boxed in black. Twelve predicted β-strands (β1–β12) are shown as arrows. Conserved blocks A, B, F, and G of the intein are indicated in bold italics.

proteins [22]. Often they are embedded into proteins involved in replication, transcription, or in related processes such as the metabolism of nucleotides [23]. According to InBase [8], 25 inteins have been detected to date in the *prp8* gene of fungi. The PRP8 protein is one of the largest and most highly conserved nuclear proteins occupying a central position in the catalytic core of the spliceosome [24]. All of the PRP8 inteins are inserted at the same position within the PRP8 sequence, but only ten of them contain endonuclease domains, while the other 15 are mini-inteins [8,23,25].

In a recent study, we investigated the distribution of PRP8 inteins in the genus *Penicillium* [26]. Among other species of this genus, *Penicillium chrysogenum*, the major producer of the β-lactam antibiotic penicillin, possesses a 157-aa mini-intein in the PRP8 protein (Fig. 1) [26–28]. We demonstrated that the *P. chrysogenum* PRP8 mini-intein undergoes autocatalytic protein splicing when heterologously expressed in a model host protein in *Escherichia coli*. The PRP8 intein of *P. chrysogenum* is thus among the smallest known nuclear-encoded active-splicing protein elements [26]. Here, we tested three split sites in the *P. chrysogenum* PRP8 intein to produce a fungal split intein capable of protein splicing in *trans*. Protein *trans*-splicing occurred at two sites: one corresponds to the insertion site of the endonuclease domain in allelic large PRP8 inteins and the other was detected at a new position, located N-terminal of the endonuclease insertion site. In the future, the split PRP8 intein of *P. chrysogenum* may be potentially useful for engineering and manipulating proteins as well as for various applications in protein chemistry.

Materials and methods

Construction of split-intein plasmids. Cloning and propagation of recombinant plasmids was done under standard conditions [29]. *Escherichia coli* strain SURE was used as the bacterial host strain for plasmid amplification [30]. The coding sequence of the *P. chrysogenum prp8*-intein with 5 N- and 4 C-terminal flanking residues (AM042015) was fused to a RGS-His-tag and cloned into the GST-vector pGEX-4T-1 (Amersham Bioscience, Europe GmbH, Freiburg, Germany) resulting in plasmid pGPch-1 [26]. To split the intein at different positions, a spacer was inserted into the intented split site by an inverse PCR procedure with pGPch-1 as a circular template. Primer pairs used were specific to the split site in diverging orientations and included overhangs to generate a spacer region. By annealing the blunt ends of the amplicon a spacer containing two stop codons at the 3′-end of the GST-intein part, a ribosome-binding site, a *Hin*dIII-restriction site and a start codon at the 5′-end of the intein-His part was created between the two intein moieties. The primer sequences are as follows: Split-N_r 5′AAGCTTTCTAGATCATTAATCCTCTCTATAAAGCACCAG, Split-N2 5′GAGCTCAAAGAGGAGAAGCTTATGGGTTCCAAGAATGTGGAGAAG, Split-E_r 5′AAGCTTTCTAGATCATTAGCCCTTCTCAGCCCTAGGAGA, Split-E2 5′GAGCTCAAAGAGGAGAAGCTTATGGCCGATGATTCGGCTCAAACG, Split-C_r 5′AAGCTTTCTAGATCATTACTCGGATTCGAGGCTAACTTG and Split-C2 5′GAGCTCAAAGAGGAGAAGCTTATGAAGACAGAGTGGGCTGGTTTC. After ligation of the linear PCR amplicons and transformation into the bacterial host strain, the resulting plasmids, pGPch-N, pGPch-E, and pGPch-C, consist of a GST-intein fragment ending with a new stop codon and a ribosome-binding site, a *Hin*dIII-restriction site and a start codon to produce the C-terminal moiety of the intein with the His-tag (Fig. 2). The spacer was inserted at three different positions: split N, split E, and split C (Fig. 1). For the PCR reactions, the Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, Inc.) was used. The DNA sequences of all plasmids were verified through DNA sequencing (sequencing service of the Department of Biochemistry, Ruhr-University Bochum, Germany).

Western blot analysis. Plasmid-encoded proteins were produced in *E. coli* strain BL21 (DE3) Gold (Stratagene, La Jolla, CA, USA). Protein production was carried out in 50 ml of Luria Broth (LB) medium at 37 °C containing ampicillin (60 μg/ml), which was inoculated from a 5 ml overnight culture. Cells were grown to a culture density of $A_{600} = 0.6$, subsequently, isopropyl-1-thio-β-D-galctopyranoside (IPTG) was added to a final concentration of 0.5 mM to induce production of recombinant proteins. After incubation for 4 h, cells were harvested by centrifugation and lysed in SDS-containing gel-loading buffer at 98 °C in a heating block. Electrophoresis was performed in a 15% SDS–polyacrylamide gel. Proteins were blotted to polyvinylidene difluoride Western blotting membrane (Biometra, Göttingen, Germany) using a semi-dry blotting system (Biometra, Göttingen, Germany). The detection was carried out with a polyclonal anti-RGS-His antibody (Qiagen, Hilden, Germany) using the chemiluminescence Western blotting kit (Roche, Germany) as described by the manufacturers.

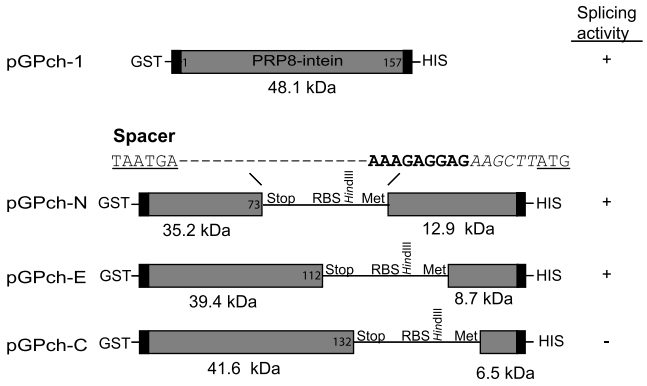

Fig. 2. Plasmid constructs used in this work. Schematic diagram depicting fusion proteins of glutathion *S*-transferase (GST), short extein sequences (black), intein and His-tag (HIS). A spacer sequence including two stop codons (underlined), a ribosome-binding site (bold), a *Hin*dIII-restriction site (italics) and a start codon (underlined) are inserted at the positions indicated in Fig. 1. Molecular weights of fusion protein precursors and products of the split fusion genes are given below. Splicing activity of the constructs is indicated aside.