The Arthrobacter Species FB24 Arth_1007 (DnaB) Intein Is a Pseudogene

Kazuo Tori, Francine B. Perler*
New England Biolabs, Inc., Ipswich, Massachusetts, United States of America

Abstract

An Arthrobacter species FB24 gene (locus tag Arth_1007) was previously annotated as a putative intein-containing DnaB helicase of phage origin (Arsp-FB24 DnaB intein). However, it is not a helicase gene because the sequence similarity is limited to inteins. In fact, the flanking exteins only contain 66 amino acids. Therefore, the intein should be referred to as the Arsp-FB24 Arth_1007 intein. The Arsp-FB24 Arth_1007 intein failed to splice in its native precursor and in a model precursor. We previously noted that the Arsp-FB24 Arth_1007 intein is the only putative Class 3 intein that is missing the catalytically essential Cys at position 4 of intein Motif F, which is one of the three defining signature residues of this class. Additionally, a catalytically essential His in position 10 of intein Motif B is also absent; this His is the most conserved residue amongst all inteins. Splicing activity was not rescued when these two catalytically important positions were ‘reverted’ back to their consensus residues. This study restores the unity of the Class 3 intein signature sequence in active inteins by demonstrating that the Arsp-FB24 Arth_1007 intein is an inactive pseudogene.

Introduction

Inteins are protein splicing elements that remove themselves from host proteins (exteins) during post-translational processing. Intein-mediated protein splicing does not require any exogenous enzymes or cofactors. Inteins are recognized as insertions within other genes and their protein products. They share four conserved splicing domain motifs (A, B, F and G, Figure 1) [1,2,3,4], with many of the most highly conserved residues playing catalytic roles [1,5,6,7]. Some inteins are chimeric proteins with a centrally located homing endonuclease domain containing four endonuclease motifs (C, D, E, and H, Figure 1) [2,3,4]. The His at position 10 of intein Motif B (B:10) is the most conserved intein amino acid (aa) and is essential for splicing in all inteins previously tested [1,5,6,8,9,10,11]; however, it is an Asn in the Arthrobacter species FB24 (Arsp-FB24) Arth_1007 (DnaB) intein. Only this intein and the Thermococcus kodakaraensis KOD1 (Tko) CDC21-1 intein (and its orthologs) lack a His at B:10 [6], bringing into question the activity of these inteins and if functional, whether they have evolved to use other residues to compensate for the lack of this catalytically essential His[8-10].

Inteins have been divided into 3 classes based on specific signature sequences and protein splicing mechanisms [1,10,11]. Most inteins are Class 1 inteins that splice themselves out of precursor proteins by a four-step mechanism [5,7,12] consisting of an initial acyl shift of the intein N-terminal Ser, Thr or Cys to form a linear (thio)ester intermediate, followed by a transesterification reaction that results in a branched intermediate (Figure 2). This branched intermediate is resolved by cyclization of the intein C-terminal Asn, which separates the intein from the ligated exteins.

A standard peptide bond is then formed between the exteins by an acyl shift. Class 2 inteins begin with other residues, but their splicing motifs are otherwise similar to Class 1 inteins [6,13]. Class 1 and Class 2 inteins splice by a mechanism that proceeds through a single branched intermediate (Figure 2) [5,7,12,13].

Class 3 inteins also lack the Class 1 N-terminal nucleophile, but have an additional class specific WCT motif consisting of a Thr at intein Motif B position 12 (B:12), a Cys at intein Motif F position 4 (F:4) and a Thr at intein Motif G position 5 (G:5) [1]. Class 3 inteins splice by a mechanism that includes two branched intermediates (Figure 2), with Cys F:4 being the nucleophile and branch point for the Class 3 specific branched intermediate [1,10,11].

The generally accepted assumption is that Class 2 and Class 3 inteins arose from mutation of the N-terminus of a Class 1 intein. To date, all experimental substitutions of a Class 1 intein N-terminus always block splicing if the change is not conservative (Ser, Thr or Cys), so these naturally mutated inteins most likely failed to splice or spliced very poorly unless and until further mutations restored robust splicing. Class 2 inteins solved this problem by overcoming the barrier to direct attack of an amide bond at the N-terminal splice junction by the C-terminal nucleophile (Fig. 2, Step 1 in Class 2 inteins) that is present in all Class 1 inteins tested. Class 3 inteins solved this problem by having Cys F:4 attack the amide bond at the intein N-terminus to form the class specific branched intermediate (Figure 2, Step 1 in Class 3 inteins), which then forms the standard Class 1 intein branched intermediate after a transesterification reaction.

To date, Class 2 inteins have only been found in KibA genes [6,13]. Class 3 inteins were previously found to be monophyletic,
while other helicase inteins, phage-derived inteins and Class 1 inteins having Cys at F:4 were polyphyletic [11]. This led to the hypothesis that all Class 3 inteins arose from a phage encoded progenitor intein that lost its N-terminal Ser, Thr or Cys [11]. Based on mutagenesis studies of modern day Class 1 inteins, these early Class 3 inteins would not splice well, if at all. They could have been retained in the population because the extein function was provided by other phage co-infecting the cell or by the host. Eventually, these early Class 3 inteins accumulated second site mutations that enabled them to splice as efficiently as standard Class 1 inteins, as exemplified by the Class 3 Mycobacteriophage Bethlehem (MP-Be) DnaB intein [1], Deinococcus radiodurans (Dra) Snf2 intein [10] and Mycobacteriophage Catera (MP-catera) Gp206 intein [11], which all spliced efficiently in a model precursor consisting of the intein flanked by the Escherichia coli Maltose Binding Protein (M) and the DSal fragment of Dirofilaria immitis paramyosin (P).

The Arsp-FB24 DnaB intein, which was annotated to be of phage origin [14], is a Class 3 intein based on phylogenetic analysis (Figure 3) and it fulfils all of the sequence criteria listed above for Class 3 inteins except that the catalytically essential CysF:4 is absent [1,10,11]. This suggests that either it is an inactive intein or it is not a Class 3 intein.

Methods
Cloning and mutagenesis
All clones were sequenced by the New England Biolabs Core facility and all enzymes were obtained from New England Biolabs (Ipswich, MA) and used as described by the manufacturer. All primers were obtained from Integrated DNA Technologies (San Diego, CA).

The Arthrobacter species FB24 (Arsp-FB24) Arth_1007 intein precursor (locus tag Arth_1007, accession number YP_830503) with an N-terminal His tag was synthesized by DNA2.0, Inc (Menlo Park, CA). Mutations were made in the homing nuclease domain active site to block endonuclease activity (Asp123Ala and Asp219Ala). The C-extein deletion mutant was constructed using a PhusionTM site-directed mutagenesis kit (New England Biolabs) with appropriate primers that truncated the precursor after the intein C-terminal Asn332.

The Arsp-FB24 Arth_1007 intein with 5 native extein residues on both sides was cloned by PCR into a model precursor termed MIP, with the intein flanked by the E. coli Maltose Binding Protein (M) and the DSal fragment of Dirofilaria immitis paramyosin (P) as previously described [1,10,11,13]. All mutations were constructed using the PhusionTM site-directed mutagenesis kit (New England Biolabs) with appropriate primers to introduce the desired mutation.

Expression, purification, and protein characterization
Precursors were expressed in either the E. coli NEB Turbo strain or NEB Express strain (New England Biolabs) by induction with 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) at OD₆₀₀ of 0.4–0.6 in 10 ml LB medium containing 100 µg/ml ampicillin for 2 hours at 37°C or 15°C overnight. Soluble lysates were boiled for 5 min in SDS Sample Buffer plus DTT (New England Biolabs), loaded onto 10–20% Tris Glycerine polyacrylamide gels.
(Invitrogen, Carlsbad, CA) and either stained with Simply Blue Safe Stain (Invitrogen) or transferred to nitrocellulose membranes for Western Blot analysis with antisera against the Maltose Binding Protein (New England Biolabs), paramyosin or the His tag (Merck, Germany) as described previously [1,10,11]. IRDye 680 secondary antibody (Li-Cor Bioscience, Lincoln, NE) was used. The membrane was scanned using an Odyssey infrared imaging system (Li-Cor Bioscience) at 700 nm.

**Phylogenetic analysis**

Bayesian inference analysis was performed using the Geneious Pro 5.1 suite of programs (Geneious, Auckland, NZ). Briefly, intein splicing domain Motifs A, B, F and G present in the InBase database [http://www.neb.com/neb/inteins.html] were concatenated to produce a single 49 aa sequence for each intein using names defined in InBase, as previously described [6,11]. Due to the variable size of Motif F, only the first and last 7 positions of Motif F were included. The concatenated sequences are listed in Tori and Perler [11]. MrBayes [15] was used with default parameters to create trees with final standard deviation of split frequencies of 0.01 or less. The data reported herein represents the Class 3 intein clade, which was found to exclude inteins from all other classes when 148 intein sequences that included all Class 3 inteins, all phage inteins, selected inteins with Cys at F:4, and selected helicase inteins were previously analyzed [11].

**Results**

**Biochemical characterization of the Arsp-FB24 Arth_1007 intein**

The Arsp-FB24 Arth_1007 intein with its complete native N-extein (52 aa) and complete native C-extein (14 aa) was expressed in E. coli at 15 or 37 °C. A single band at the predicted size of the precursor (NIC, 44 kDa) was observed in the soluble protein fraction by both staining with Simply Blue Stain or Western Blot with antisera directed against an N-terminal His tag and an N-terminal His tag (Figure 4 and data not shown). No spliced product or free intein were observed. As a control, soluble protein samples from a clone expressing an experimentally generated C-terminal cleavage product consisting of the same N-extein and the intein (NI, 42.5 kDa) were co-electrophoresed along with the NIC precursor. NI migrated faster than the complete NIC precursor (Figure 4). No N-terminal (6.4 and 37.7 kDa) or C-terminal (42.5 and 1.5 kDa) single splice junction cleavage products were observed in the NIC precursor sample and no N-terminal cleavage products were observed in the truncated NI precursor sample.

Further characterization was performed on a model precursor (MIP) that has been used to study splicing of numerous inteins because it allows facile identification of splicing and cleavage products that can be clearly differentiated based on relative mobility in SDS-PAGE or immunoreactivity [1,10,11,13]. The Arsp-FB24 Arth_1007 intein with 5 native extein residues on both sides (I) was cloned.
between the E. coli Maltose Binding Protein (M) and a Paramyosin fragment (P) to generate the Arsp-FB24 Arth_1007 intein MIP precursor. This precursor failed to splice or yield off-pathway cleavage products in vivo when expressed at 15, 30 or 37°C, or when incubated in vitro at 37°C (Figure 5 and data not shown).

Mutations that restore missing conserved amino acids

Since His\textsuperscript{B:10} is essential for splicing in all inteins tested, Asn\textsuperscript{B:10} (B:10) was ‘reverted’ back to His. Again, no activity was detected in vivo at 15, 30 or 37°C, or after incubation in vitro at 37°C (Figure 5 and data not shown).

As a putative Class 3 intein, a Cys at position F:4 in the Arsp-FB24 Arth_1007 intein would be required for splicing [1,10,11]. The F:4 position in the Arsp-FB24 Arth_1007 intein was originally assigned as Asp\textsuperscript{315} [6]. No activity was rescued when Asp\textsuperscript{315} was mutated to Cys alone or in combination with Asn\textsuperscript{B:10}His (Figure 5).

The InBase intein-specific BLAST tool (http://www.neb.com/neb/inteins.html) yielded the highest similarity scores between the Arsp-FB24 Arth_1007 intein versus the Mycobacteriophage...
Mutated to Cys or when the Asn65His and Gly311Cys mutations were combined (Figure 5 and data not shown).

However, no activity was observed when Gly 311 was mutated to Cys or when the Asn65His and Gly311Cys mutations were combined (Figure 5 and data not shown).

Discussion

The Arsp-FB24 Arth_1007 intein failed to splice or perform single splice junction cleavage reactions under any condition tested in its native precursor or in a model precursor, even after essential Class 3 intein catalytic residues were mutated back to the consensus amino acid at position B:10 (Asn⁶⁵) or possible F:4 positions (Asp₃₁₅ or Gly₃¹¹). Simply Blue Safe stained SDS-PAGE of soluble lysates after in vivo expression at 37 °C and after incubation in vitro at 37 °C at the indicated pH values in the presence (+) or absence (−) of 50 mM DTT. Western Blots with anti-P sera failed to detect spliced MP or cleavage products in all samples (data not shown). Molecular weight standards are in the left lane of each gel and selected sizes are listed in kDa (New England Biolabs 10 to 250 kDa ladder).

doi:10.1371/journal.pone.0026361.g004

Omega (MP-Omega) DnaB intein (38% aa identity and 51% aa similarity) and the Mycobacteriophage Catera (MP-Catera) Gp206 intein (35% aa identity and 51% aa similarity) [6,11].

Based on a pairwise sequence alignment to both of these inteins using both the InBase BLAST tool and BLASTP on the NCBI website (http://blast.ncbi.nlm.nih.gov), the Arsp-FB24 Arth_1007 intein position F:4 was reassigned as Gly₃¹¹ (Figure 1 and data not shown).

However, no activity was observed when Gly₃¹¹ was mutated to Cys or when the Asn⁶⁵His and Gly₃¹¹Cys mutations were combined (Figure 5 and data not shown).

No previous examples of decayed inteins have been reported, probably because most inteins are found in essential proteins [3,6,16] whose loss would result in reduced viability or cell death. If present in a non-essential intein position, inactivation of such an intein would not be recognizable with time because genetic drift would obliterate intein signatures. The Arsp-FB24 Arth_1007 intein may have been retained in the Arthrobacter species FB24 genome because the prophage genome remnant is irrelevant to survival of the host cell. Alternatively, it may have only recently become inactive and is in the process of being deleted. Another intriguing possibility exists. We previously proposed that Class 3 inteins arose from a single standard phage intein that lost its N-terminal nucleophile [11]. The progenitor Class 3 intein could persist in a phage population until second-site mutations restored splicing efficiency if the intein function was provided by the host cell or by other phage. The Arsp-FB24 Arth_1007 intein could represent a failed intermediate in such an evolutionary progression from the progenitor Class 3 intein.

By demonstrating that the Arsp-FB24 Arth_1007 intein is not functional under any conditions tested, this study resolves the
confusion caused by the identification of the Arsp-FB24 Arth_1007 intein as a Class 3 intein based on phylogenetic analysis (Figure 3) and the presence of some Class 3 signature sequence components (Trp at B:12 and Thr at G:5) [1], despite the absence of Cys at position F:4, which performs the mechanistic step that defines Class 3 inteins (Figure 2).

Acknowledgments
We thank Manoj Cheriyan (New England Biolabs) for helpful discussions, Lise Raleigh for reading of the manuscript (New England Biolabs) and Don Comb (New England Biolabs) for support and encouragement.

Author Contributions
Conceived and designed the experiments: KT FBP. Performed the experiments: KT. Analyzed the data: KT FBP. Wrote the paper: KT FBP.

References
1. Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, et al. (2010) Splicing of the Mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate Block F nucleophile. J Biol Chem 285: 2515–2526.
2. Perler FB, Olsen GJ, Adam E (1997) Compilation and analysis of intein sequences. Nucleic Acids Res 25: 1087–1093.
3. Pietrokovski S (1998) Modular organization of inteins and C-terminal autocatalytic domains. Protein Sci 7: 64–71.
4. Pietrokovski S (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. Protein Sci 3: 2140–2150.
5. Mills KV, Perler FB (2005) The mechanism of intein-mediated protein splicing: variations on a theme. Protein Pept Lett 12: 751–755.
6. Perler FB (2002) InBase: the Intein Database. Nucleic Acids Res 30: 383–384.
7. Paulus H (1998) The chemical basis of protein splicing. Chemical Society Reviews 27: 375–386.
8. Kawasaki M, Nogami S, Satow Y, Ohya Y, Anraku Y (1997) Identification of three core regions essential for protein splicing of the yeast Vma1 endonuclease. J Biol Chem 272: 15668–15674.
9. Romanelli A, Shekhman A, Coombes D, Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. Proc Natl Acad Sci U S A 101: 6397–6402.
10. Brace LE, Southworth MW, Tori K, Cushing ML, Perler F (2010) The Deinococcus radiodurans SmfZ intein caught in the act: detection of the Class 3 intein signature Block F branched intermediate. Protein Sci 19: 1525–1533.
11. Tori K, Perler FB (2013) Expanding the definition of class 3 inteins and their proposed phage origin. J Bacteriol 193: 2035–2041.
12. Noren CJ, Wang J, Perler FB (2000) Dissecting the Chemistry of Protein Splicing and Its Applications. Angew Chem Int Ed Eng 39: 450–466.
13. Southworth MW, Benner J, Perler FB (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. Embo J 19: 5019–5026.
14. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, et al. (2009) The integrated microbial genomes system: an expanding comparative analysis resource. Nucleic Acids Res 38: D382–390.
15. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294: 2310–2314.
16. Dalgaard JZ, Moser MJ, Hughey R, Mian IS (1997) Statistical modeling, phylogenetic analysis and structure prediction of a protein splicing domain common to inteins and hedgehog proteins. J Comput Biol 4: 193–214.