Available online at www.sciencedirect.com
ScienceDirect
Biochemical and Biophysical Research Communications 366 (2008) 239–243
BBRC
www.elsevier.com/locate/ybbrc

Minimization of a eukaryotic mini-intein

Skander Elleuche, Kristin Döring, Stefanie Pöggeler *

Institute of Microbiology and Genetics, Department of Genetics of Eukaryotic Microorganisms, Georg-August-University of Göttingen, Grisebachstrasse 8, 37077 Gottingen, Germany

Received 5 November 2007
Available online 3 December 2007

Abstract

Inteins are internal protein splicing elements that can autocatalytically self-excise from their host protein and ligate the protein flanks (exteins) with a peptide bond. Large inteins comprise independent protein splicing and endonuclease domains whereas mini-inteins lack the central endonuclease domain. To identify mini-intein domains that are essential for protein splicing, deletions were introduced at different sites of the 157-aa PRP8 mini-intein of *Penicillium chrysogenum*. The removal of eight and six amino acids at two different sites resulted in a functional eukaryotic mini-intein of only 143 aa.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Intein; Protein splicing; *Penicillium chrysogenum*; PRP8 protein; Minimization

An intein is a self-catalytic sequence that is embedded in-frame within a precursor protein. To produce a mature protein the intein precisely self-excises from the host protein. It thereby ligates the flanking sequences (N- and C-exteins) [1,2]. The post-translational removal of the intein is termed protein splicing and depends on three highly conserved residues. Most inteins begin with a serine, cysteine, or threonine residue at the N terminus, and end with an asparagine residue at the C terminus. The first position of the C-terminal extein is either a hydroxyl-containing (serine) or a thiol-containing (cysteine) residue. Protein splicing typically involves four steps: (i) an N-O or N-S acyl shift at the N-extein/intein junction; (ii) a transesterification which transfers the N-extein to the side-chain of the first residue (+1) of the C-extein; (iii) cyclization of a conserved asparagine residue at the C terminus of the intein and cleavage of the peptide bond, resulting in the release of the intein; and (iv) rearrangement to a peptide bond of the ester/thioester bond linking the N- and C-exteins [3–6].

Inteins defined as being large or minimal (mini-) depending on whether or not they contain an endonuclease domain [2]. The splicing domain of large inteins seems to be split by the endonuclease domain into an N-terminal and a C-terminal subdomain. Because several splicing-efficient mini-inteins have been engineered from large inteins by deleting the central endonuclease domain, it is obvious that the endonuclease domain is not involved in protein splicing [7–10]. Both intein subdomains contain conserved blocks of amino acids (blocks A, N2, B, and N4 for the N-terminal subdomain and blocks G and F for the C-terminal subdomain) [11–14]. These domains can also be identified in mini-inteins. Three-dimensional structures of inteins reveal that the N- and C-terminal splicing domains form a common horseshoe-like 12-β-strand scaffold termed the Hedgehog/intein (Hint) module [15–18].

Most inteins have been found in the genomes of prokaryotes. In eukaryotes, they have been identified in ~50 taxa, encoded within the nuclear genomes of fungi and the plastomes of some unicellular algae [12]. In fungi, most of the inteins (13 large and 19 mini-inteins) have been detected in the *prp8* gene [12,19]. 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 [20].

In a previous study [21], we identified that the PRP8 protein of *Penicillium chrysogenum* (*Pch*), the major pro-

* 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.2007.11.126

ducer of the β-lactam antibiotic penicillin, possesses a 157-aa mini-intein in the PRP8 protein. The *Pch* PRP8 mini-intein can undergo autocatalytic protein splicing when heterologously expressed in a model host protein in *Escherichia coli* [21]. Thus, the *Pch* PRP8 intein is among the smallest known nuclear-encoded active splicing protein elements. Moreover, we recently demonstrated that the *Pch* intein is capable of protein splicing in *trans* [22]. Several studies of the sequence and structural requirements of protein splicing have been conducted for large inteins [7–10]. However, little is known about the catalytic and structural elements that participate in protein splicing of naturally occurring mini-inteins. Here, we report the minimization of the eukaryotic *Pch* PRP8 mini-intein. Amino acid sequences of this intein can be deleted at two different sites without affecting splicing activity. One site corresponds to the insertion site of the endonuclease domain in large allelic PRP8 inteins. The other site was detected at a new position corresponding to the insertion site of a putative tongs domain of a large fungal PRP8 intein. The smallest functional intein found in our study comprises only 143 residues and is the smallest functional eukaryotic intein engineered so far.

Materials and methods

Sequence analysis. Protein sequence alignments were performed either using the ClustalX program [23] or the LALIGN program [24]. Protein secondary structure prediction was done with the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/) [25].

Construction of truncated P. chrysogenum inteins. Escherichia coli strain SURE (Stratagene, La Jolla, CA, USA) was used for general plasmid construction. Cloning and propagation was done under standard conditions [26]. Deletion of selected regions of the *Pch* PRP8 intein (AM042015) was done by inverse PCR using plasmid pGPch-1 as a circular template [21,27]. Primer pair Endo_for (CAAACGCACAGTTTCAAGA TTGAGCAAGTTAGCCTCG) and Endo_rev (CTTCTCAGCCCGGG GAGATGTAAAGGC, SmaI restriction site underlined) were used to construct plasmid pΔE. The SmaI restriction site in primer Endo_rev was used for the selection of positive clones. Amplification with primer pair TSD_for (CAAACGGTGGAGATCACTGCTGCCGAGTTTGCCGC GCTTTCTACC) and TSD_rev (CTCTCTATAAAGCACCAGAATA TGGTTCGGCGTCACCACAAGGTCC) resulted in plasmid pΔT and with TSD_for and TSD2_rev (CACCAGAATATGGTTCGGCGTCAC CACAAGGTCC) in plasmid pΔTΔ4. To construct plasmid pΔT-E, TSD_rev and Endo_for were used as primer pair. Plasmid pΔTΔE was generated by means of inverse PCR with plasmid pΔE as template and primers TSD_for and TSD-rev. Inverse PCRs were performed using Phusion™ High-Fidelity DNA Polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany). Subsequently, amplicons were treated with Klenow Polymerase (Fermentas, St. Leon-Rot, Germany) to generate blunt-ended products and with polynucleotide kinase (Fermentas, St. Leon-Rot, Germany) to generate phosphorylated 5′-ends. After ligation of the linear PCR-fragments, plasmids were transformed into the bacterial host strain. Positive clones were screened by colony PCR [28] using MolTaq (Molzym GmbH & Co., KG, Bremen, Germany). The DNA sequences of all plasmids were verified by DNA sequencing at the G2L-sequencing service of the Göttinger Genom Labor (Georg-August University, Göttingen, Germany).

Protein expression in E. coli. Plasmid encoded intein constructs were overexpressed in *E. coli* strain BL21 (DE3) Gold (Stratagene, La Jolla, CA, USA) as described earlier [21]. In brief, 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 under constant shaking at 37 °C 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.

SDS–polyacrylamide gel electrophoresis (PAGE) and Western blotting analysis. Proteins were separated in a 15% SDS–polyacrylamide gel and transferred to a polyvinylidene difluoride membrane (Biometra, Göttingen, Germany) by using a semidry blotting system (Biometra, Göttingen, Germany). Western blots were either done with a polyclonal anti-His-RGS antibody (Qiagen, Hilden, Germany) or with an anti-glutathione S-transferase (GST) HRP-conjugate (Amersham Bioscience, Europe GmbH, Freiburg, Germany). Signal detection was carried out using the chemiluminescence Western blotting kit (Roche, Mannheim, Germany) according to the supplier's recommendations.

Results and discussion

Selection of deletion sites

Little is known about the catalytic and structural elements that participate in the protein splicing of eukaryotic mini-inteins. As a guide to constructing deletion mutations, the *Pch* PRP8 mini-intein was aligned with the related large PRP8 intein of *Aspergillus fumigatus* (*Afu*) which is known to splice in a recombinant protein in *E. coli* [29]. *Afu* PRP8 intein is unusual because, in addition to a 454-aa endonuclease domain, it has a 222-aa extra sequence which is predicted to be a putative tongs subdomain [29]. This prediction was based solely on the location of this domain being the same as the tongs subdomain in the crystal structure of vacuolar ATPase subunit (VMA) intein of the yeast *Saccharomyces cerevisiae* (*Sce*). In the *Sce* VMA, the tongs domain is supposed to participate in DNA substrate binding [30]. Alignment of the *Pch* PRP8 mini-intein and the *Afu* PRP8 large intein without the tongs and endonuclease domains revealed that the proteins share a high degree of amino acid identity within the first 72 N-terminal residues and 40 C-terminal residues. By contrast, the middle parts displayed a low level of sequence conservation. Conserved motifs A, N2, and B were identified within the conserved N-terminal subdomains and motifs F and G within the conserved C termini (Fig. 1). The N4 motif, comprising 16 residues and including a highly conserved Asp or Glu residue [13] could not be allocated in either the *Pch* or the *Afu* PRP8 intein. The dependence of protein splicing on conserved residues within motifs A, B, F, and G has recently been assessed for the PRP8 mini-intein of the basidiomycete *Cryptococcus neoformans* [31]. Thus, a deletion mutation (pΔT-E, a deletion of 45 aa) was constructed so that its deleted area matched the poorly conserved region between the *Afu* and the *Pch* inteins. The protein expressed from pΔT-E was 112 aa and comprised only the conserved N- and C-terminal subdomains (Fig. 2).

To investigate amino acid residues surrounding putative tongs and/or endonuclease domain insertion sites, we constructed the plasmids pΔT (deletion of 8 aa), pΔE (deletion of 6 aa), and pΔTΔE (deletion of 8 aa and 6 aa). We took

Fig. 1. Alignment of PRP8 inteins from *Aspergillus fumigatus* and *Penicillium chrysogenum.* Extein sequences are boxed in black. Conserved intein blocks A, N2, B, F, and G are shown in bold italics. Deleted amino acids of cloned intein variants are indicated in white and shaded in gray. Positions of a putative tongs domain (TSD) and the endonuclease (ENDO) domain of *A. fumigatus* are given. Split sites are indicated as arrows. Abbreviations and accession numbers are as follows: *Afu*, *Aspergillus fumigatus* (AAV91021) and *Pch*, *Penicillium chrysogenum* (AM042015). Gaps introduced to optimize alignment; (:), mark identical amino acids, (.), mark similar amino acids.

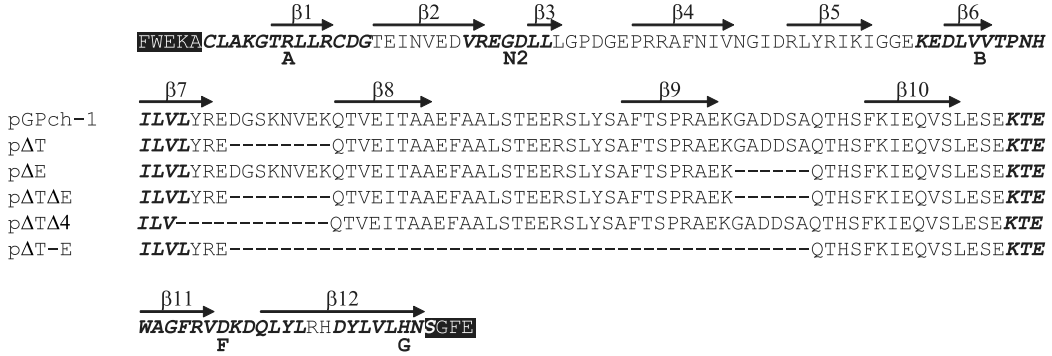

Fig. 2. Amino acid sequence of the 157-aa *P. chrysogenum* PRP8 mini-intein (AM042015) and of deletion mutants. Extein sequences and conserved blocks are marked as in Fig. 1. Twelve predicted β-strands (β1–β12) are indicated as arrows. Deletions within different intein constructs are indicated as gaps.

into account that the functionality of the PRP8 inteins might depend on the integrity of β-strands within the intein structure. Similar to the known horseshoe-like Hint domain structure of the engineered DnaB mini-intein of the cyanobacterium *Synechocystis* sp. (*Ssp*) [18], the *Pch* PRP8 intein is predicted to possess 12 β-strands (Fig. 2). Therefore, the plasmids pΔT, pΔE, and pΔTΔE carry deletions that should not affect the predicted β-strand structure. However, the deletion of 12 aa from the construct pΔTΔ4 (four additional amino acids deleted, in comparison to pΔT) should lead to a truncation of the predicted β7 strand (Fig. 2).

Effect of deletions on protein splicing of the Pch PRP8 mini-intein

The *cis*-splicing activity of the *Pch* PRP8 intein has been tested and confirmed previously by constructing plasmid pGPch-1, in which *Pch* PRP8 (plus a 5-aa and 4-aa native N-extein and C-extein sequence, respectively) was fused to an N-terminal GST protein and a C-terminal His-tag [21]. We used this plasmid and generated the desired derivatives via an inverse PCR method and subsequent ligation of the obtained amplicons. The resulting plasmids encoded intein derivatives that ranged in size from 112 aa to 151 aa (Fig. 3A). The expression of the mutated tripartite fusion genes was studied in *E. coli* in parallel with the parent plasmid pGPch-1, which has an intact intein-coding sequence. Overexpression of the fusion protein was induced by the addition of IPTG, and the precursor and products of the splicing reaction were subjected to SDS–PAGE. For derivatives encoded by pΔT, pΔE, and pΔTΔE the splice products, the ligated exteins (GST-His, 32.2 kDa) and the free intein (~17 kDa) were readily detected on Coomassie gel (Fig. 3B). With the exception of the full-length intein (pGPch-1) and the mutant derivative encoded by pΔE, pre-

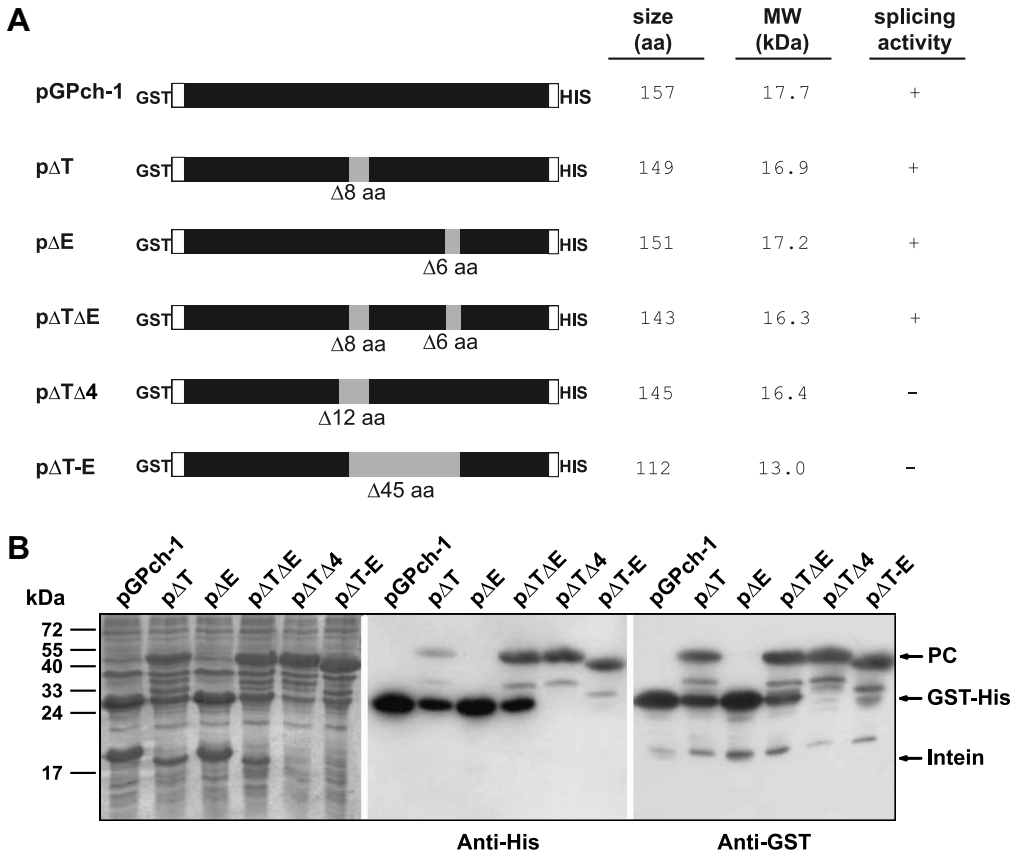

Fig. 3. *Cis*-splicing activity of the *P. chrysogenum* PRP8 mini-intein containing deletions. (A) Schematic diagram of plasmid constructs depicting fusion proteins of the glutathion S-transferase (GST), N-extein (5 aa), and C-extein (4 aa) (whites boxes), intein (black), and His-tag (His). Deleted regions are shaded in gray at the appropriate position. Size in amino acids (aa), molecular weight (MW) and splicing activity of the corresponding intein is indicated aside. (B) SDS–PAGE and Western blot of lysates from *E. coli* expressing different variants of the *P. chrysogenum* intein fusion proteins. Total cellular proteins of *E. coli* strains containing the depicted plasmids were resolved on SDS–PAGE and visualized by Coomassie blue staining or Western blotting analysis using anti-His RGS-antibody or anti-GST HRP-conjugate. Positions of the spliced product (GST-His, 32.2 kDa) as well as precursor proteins (PC) are indicated.

cursors were also clearly visible. Importantly, there was accordance between the predicted and the observed size of the different inteins, whereas the size of the splice product was constant for all derivatives. Western blot analysis using anti-His and anti-GST antisera confirmed the identification of spliced products and precursors. In addition to the precursor, N-terminal degradation products of the precursor containing the His-tag and parts of the GST-tag were detected in mutant derivatives encoded by pΔT, pΔTΔE, pΔTΔ4, and pΔT−E. A signal of ~20 kDa that was detectable in the anti-GST Western blot in every single line was probably caused by an unspecific reaction of the GST antibody. The protein splicing activity of derivative pΔE was indistinguishable from that of the full-length intein (pGch-1), and pΔT and pΔTΔE showed efficient splicing whereas no protein-splicing reaction was detectable in derivatives pΔT−E and pΔTΔ4. Based on our deletion study, the functional deletion derivatives (pΔT, 149 aa; pΔE, 151 aa; and pΔTΔE, 143 aa) possess the entire horseshoe-like 12-β-strand core present in the three-dimensional structures of all *cis*- and *trans*-splicing inteins analyzed so far [17]. By contrast, mutations destroying this structure resulted in the splice-deficient constructs pΔTΔ4 and pΔT−E. The smallest functional intein found in this study comprises 143 residues and is the smallest functional eukaryotic intein discovered so far. According to InBase [12], only the 138-aa *Wiv* RIR1 mini-intein identified in the eukaryotic *Wiseana iridescent* virus is smaller in size, although the splicing activity of this intein has not been proven experimentally. Similarly, mini-inteins comprising less than 143 aa have also been detected in the other two domains of life: the cyanobacterium *Spirulina platensis* contains a 136-aa mini-intein (*Spl* DnaX) and the archaeon *Thermococcus kodakaraensis KOD1* (*Tko* CDC21-1) contains a mini-intein of 140 aa. Again, however, the splicing activity of these mini-inteins has not yet been experimentally verified [12].

The smallest functional mini-intein obtained so far was engineered from the *Mycobacterium tuberculosis* (*Mtu*) recA large intein. The engineered *Mtu* recA comprises 137 residues but exhibits only a very low splicing activity [8,32]. Therefore, it has been speculated that 137 residues could be the lower limit for full protein-splicing activity [32]. By contrast, the 143-aa *Pch* PRP8 engineered in this

study exhibited an efficient splicing activity. Interestingly, this *Pch* PRP8 intein tolerates deletion not only at the putative endonuclease integration site but also at an additional site; the integration site of a putative tongs domain. These two sites correspond to the previously identified artificial split sites at D73 and G112 which enabled effective *trans*-splicing of the *Pch* PRP8 intein (Fig. 1) [22].

Thus, the *Pch* PRP8 intein is remarkably efficient at *cis*- and *trans*-splicing in heterologous systems and, furthermore, tolerates deletions at two different sites. This indicates that the *Pch* PRP8 mini-intein might have a more robust structure or folding ability than do many other inteins and that it might be of practical importance to engineer this intein for various applications.

Acknowledgment

The authors thank Dr. Nicole Nolting for critically reading the manuscript.

References

[1] F.B. Perler, E.O. Davis, G.E. Dean, F.S. Gimble, W.E. Jack, N. Neff, C.J. Noren, J. Thorner, M. Belfort, Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature, Nucleic Acids Res. 22 (1994) 1125–1127.
[2] X.Q. Liu, Protein-splicing intein: genetic mobility, origin, and evolution, Annu. Rev. Genet. 34 (2000) 61–76.
[3] C.J. Noren, J. Wang, F.B. Perler, Dissecting the chemistry of protein splicing and its applications, Angew. Chem. Int. Ed. Engl. 39 (2000) 450–466.
[4] L. Saleh, F.B. Perler, Protein splicing in *cis* and in *trans*, Chem. Rec. 6 (2006) 183–193.
[5] P.L. Starokadomsky, Protein splicing, Mol. Biol. (Mosk) 41 (2007) 278–293.
[6] H. Paulus, Protein splicing and related forms of protein autoprocessing, Annu. Rev. Biochem. 69 (2000) 447–496.
[7] H. Wu, Z. Xu, X.Q. Liu, Protein trans-splicing and functional mini-inteins of a cyanobacterial *dnaB* intein, Biochim. Biophys. Acta 1387 (1998) 422–432.
[8] V. Derbyshire, D.W. Wood, W. Wu, J.T. Dansereau, J.Z. Dalgaard, M. Belfort, Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution, Proc. Natl. Acad. Sci. USA 94 (1997) 11466–11471.
[9] K. Shingledecker, S.Q. Jiang, H. Paulus, Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments, Gene 207 (1998) 187–195.
[10] S. Chong, Y. Shao, H. Paulus, J. Benner, F.B. Perler, M.Q. Xu, Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an *in vitro* splicing system, J. Biol. Chem. 271 (1996) 22159–22168.
[11] S. Pietrokovski, Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins, Protein Sci. 3 (1994) 2340–2350.
[12] F.B. Perler, InBase: the intein database, Nucleic Acids Res. 30 (2002) 383–384.
[13] S. Pietrokovski, Modular organization of inteins and C-terminal autocatalytic domains, Protein Sci. 7 (1998) 64–71.
[14] F.B. Perler, G.J. Olsen, E. Adam, Compilation and analysis of intein sequences, Nucleic Acids Res. 25 (1997) 1087–1093.
[15] T. Klabunde, S. Sharma, A. Telenti, W.R.J. Jacobs, J.C. Sacchettini, Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing, Nat. Struct. Biol. 5 (1998) 31–36.
[16] T.M. Hall, J.A. Porter, K.E. Young, E.V. Koonin, P.A. Beachy, D.J. Leahy, Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins, Cell 91 (1997) 85–97.
[17] P. Sun, S. Ye, S. Ferrandon, T.C. Evans, M.Q. Xu, Z. Rao, 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 (2005) 1093–1105.
[18] Y. Ding, M.Q. Xu, I. Ghosh, X. Chen, S. Ferrandon, G. Lesage, Z. Rao, 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 (2003) 39133–39142.
[19] M.I. Butler, J. Gray, T.J. Goodwin, R.T. Poulter, The distribution and evolutionary history of the PRP8 intein, BMC Evol. Biol. 6 (2006) 42.
[20] R.J. Grainger, J.D. Beggs, Prp8 protein: at the heart of the spliceosome, RNA 11 (2005) 533–557.
[21] S. Elleuche, N. Nolting, S. Pöggeler, Protein splicing of PRP8 mini-inteins from species of the genus Penicillium, Appl. Microbiol. Biotechnol. 72 (2006) 959–967.
[22] S. Elleuche, S. Pöggeler, *Trans*-splicing of an artificially split fungal mini-intein, Biochem. Biophys. Res. Commun. 355 (2007) 830–834.
[23] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.
[24] X. Huang, W. Miller, A time-efficient, linear-space local similarity algorithm, Adv. Appl. Math. 12 (1991) 337–357.
[25] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (2000) 404–405.
[26] J. Sambrock, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory, Cold Spring Habor, NY, 2001.
[27] T. Triglia, M.G. Peterson, D.J. Kemp, A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences, Nucleic Acids Res. 16 (1988) 8186.
[28] M.A. Hofmann, D.A. Brian, Sequencing PCR DNA amplified directly from a bacterial colony, Biotechniques 11 (1991) 30–31.
[29] X.Q. Liu, J. Yang, Prp8 intein in fungal pathogens: target for potential antifungal drugs, FEBS Lett. 572 (2004) 46–50.
[30] E. Werner, W. Wende, A. Pingoud, U. Heinemann, High resolution crystal structure of domain I of the *Saccharomyces cerevisiae* homing endonuclease PI-SceI, Nucleic Acids Res. 30 (2002) 3962–3971.
[31] E.J. Pearl, J.D. Tyndall, R.T. Poulter, S.M. Wilbanks, Sequence requirements for splicing by the *Cne* PRP8 intein, FEBS Lett. 581 (2007) 3000–3004.
[32] K. Hiraga, V. Derbyshire, J.T. Dansereau, P. Van Roey, M. Belfort, Minimization and stabilization of the *Mycobacterium tuberculosis recA* intein, J. Mol. Biol. 354 (2005) 916–926.