APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY

Skander Elleuche · Nicole Nolting · Stefanie Pöggeler

Protein splicing of PRP8 mini-inteins from species of the genus *Penicillium*

Abstract Inteins are protein-intervening sequences found inside the coding region of different host proteins and are translated in-frame with them. They can self-excise through protein splicing, which ligates the host protein flanks with a peptide bond. In this study, four different species of the genus *Penicillium* were investigated for the presence of inteins inside the conserved splicing-factor protein PRP8. We identified 157 to 162 amino acid in-frame insertions in the PRP8 protein of *Penicillium chrysogenum*, *Penicillium expansum*, and *Penicillium vulpinum* (formerly *Penicillium claviforme*). The *Penicillium* PRP8 inteins are mini-inteins without a conserved endonuclease domain. We demonstrated that the PRP8 mini-inteins of *P. chrysogenum*, *P. expansum*, and *P. vulpinum* undergo autocatalytic protein splicing when heterologously expressed in *E. coli*, in a model host protein, and in a divided GFP model system. They are, thus, among the smallest known nuclear-encoded, active splicing protein elements. The GFP assay should be valuable as a screening system for protein splicing inhibitors as potential antimycotic agents and as tools for studying the mechanism of protein splicing of fungal mini-inteins.

Introduction

Inteins are self-catalytic protein-splicing elements that disrupt a host gene and its protein product (Liu 2000). At the protein level, the intein is excised precisely and its flanking sequences, termed N- and C-exteins, are joined with a peptide bond to produce the mature spliced protein (Perler et al. 1994). Protein splicing is a posttranslational processing event that releases an internal protein sequence from a protein precursor. The mechanism of protein splicing typically consists of four steps: two acyl

S. Elleuche · N. Nolting · S. Pöggeler (✉)
Lehrstuhl für Allgemeine und Molekulare Botanik,
Ruhr-Universität Bochum,
44780 Bochum, Germany
e-mail: stefanie.poeggeler@rub.de
Tel.: +49-234-3224264
Fax: +49-234-3214184

rearrangements at the two splicing junctions, a transesterification between the two junctions, and cyclization of a residue at the C-terminal junction. The splicing reaction is, thus, dependent on three highly conserved residues at both splice-site junctions. The vast majority of inteins posses an asparagine residue at their C-termini and a hydroxyl- (serine) or thiol- (cysteine) containing residue after each splice junction (Paulus 2000; Pietrokovski 2001; Gogarten et al. 2002).

Protein-splicing elements were first described in fungi, when an in-frame insertion was identified in the *Saccharomyces cerevisiae VMA* gene encoding a vacuolar ATPase. The insertion was unrelated to the sequence of homologue ATPases and it was demonstrated that the mature VMA protein lacked the protein insertion (Hirata et al. 1990; Kane et al. 1990). Meanwhile, inteins have been found in all three domains of life: Eukaryotes, Bacteria and Archaea [InBase, the New England Biolabs Intein database (Perler 2002)]. They are present in proteins with diverse functions including metabolic enzymes, enzymes involved in DNA repair, DNA and RNA polymerases, spliceosomal factors, proteases, ribonucleotide reductases, and the vacuolar-type ATPase (Pietrokovski 2001; Perler 2002). The majority of inteins contain two discrete functional domains, the homing endonuclease domain and the splicing domain (Liu 2000; Pietrokovski 2001).

Because of their endonuclease activity, fungal inteins appear to spread by the same mechanism as that described for introns that harbour a homing endonuclease (Gogarten et al. 2002). The endonuclease domains of inteins are not involved in protein splicing (Belfort and Roberts 1997). This splicing is carried out by a ~140-amino-acid splicing domain. Inteins containing an endonuclease domain are termed large inteins; inteins without an endonuclease domain are termed minimal inteins (mini-inteins) (Liu 2000).

So far, several inteins have been detected in fungal organisms (Pöggeler 2005). They are located in the *VMA* gene of ascomycetous yeasts and in the *prp8* gene of basidiomycetes and filamentous ascomycetes. 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 (Grainger and Beggs 2005). A short intein without an endonuclease domain was shown to be present in the *prp8* gene of the fungal pathogen *Cryptococcus neoformans* (Butler et al. 2001). Subsequent studies have found allelic (occupying the same genetic site in different species) PRP8 inteins with a large endonuclease domain in the ascomycetes *Histoplasma capsulatum*, *Aspergillus nidulans*, and *Aspergillus fumigatus* (Liu and Yang 2004). Recently, large inteins have been identified in the glutamate synthase gene *glt1* of the yeasts *Debaryomyces hansenii* and *Pichia guillermondii*, as well as in the *glt1* gene and the chitin synthase gene *chs2* of the filamentous ascomycete *Podospora anserina* (Butler et al. 2005).

Inteins are typically found inserted in regions of their host proteins that are especially well-conserved across species, thus facilitating the amplification of this region by PCR. In this study, we investigated the distribution and evolutionary history of PRP8 inteins in the genus *Penicillium*. Four different species of *Penicillium* were examined: *P. expansum*, *P. chrysogenum*, *P. vulpinum* (formerly *P. claviforme*), and *P. thomii*. These species were selected because of their frequency in nature and their ability to grow on a wide variety of substrates. Moreover, some species of the genus *Penicillium* produce important secondary metabolites that are mycotoxins and antibiotics. *P. chrysogenum* is one of the major producers of the β-lactam antibiotic penicillin (Elander 2003). *P. expansum* is the causal agent of blue mold rot, the most devastating pathogen in harvested fruits. It causes significant loss to the fruit industry and is of concern in fruit products because of its production of the mycotoxin patulin (Pitt 1979; Majerus and Kapp 2002). *P. thomii* was found to have no PRP8 intein; thus, in this paper, we report the characterization of PRP8 mini-inteins of *P. chrysogenum*, *P. expansum*, and *P. vulpinum* by sequence analysis and splicing activity.

Materials and methods

Gene cloning and sequence analysis

Escherichia coli strain SURE was used as host for plasmid amplification (Greener 1990). Cloning and propagation of recombinant plasmids was done under standard conditions (Sambrook et al. 2001). *Penicillium* strains used in this study were obtained from the culture collection of the Ruhr-University Bochum, Department of General and Molecular Botany: *P. chrysogenum*, Pch-1; *P. expansum*, Pex-1; *P. vulpinum*, Pvu-1, and *P. thomii*, Pth-1. For isolation of total genomic DNA, mycelia were cultivated for 2–3 days in Petri dishes containing 20 ml of CCM (Walz and Kück 1995). The extraction of DNA was performed according to Pöggeler et al. (1997). Coding sequences of *prp8* inteins and parts of their flanking exteins were amplified by doing polymerase chain reaction using the primer pair prp8-5 (5′-ATGAA GAGCAATCCATTCTGGTGGAC-3′) and prp8-3 (5′-

TTGGCACGGTTGATGGTAGGTGACCACCA-3′). Primers were designed according to the *prp8* sequences of the fully sequenced genomes of *Aspergillus fumigatus* (at http://www.tigr.org/tdb/e2k1/afu1/), *Aspergillus nidulans*, *Fusarium graminearum*, *Magnaporthe grisea*, and *Neurospora crassa* (at http://www.broad.mit.edu/annotation/fungi).

The amplified DNAs were cloned into a pDrive plasmid vector (Qiagen, Hilden, Germany). At least two clones were subjected to automated sequencing by MWG-Biotech Customer Service (Ebersberg, Germany). Nucleotide sequences were deposited in the EMBL database under the following accession numbers: *P. thomii*, AM042014; *P. chrysogenum*, AM042015; *P. vulpinum*, AM042016; and *P. expansum*, AM042017.

Sequence analysis

Protein sequence alignments were performed using the ClustalX program (Thompson et al. 1997).

Protein splicing analysis in *E. coli*

To construct gene expression plasmids, PRP8 intein coding sequences with 5 N- and 4 C-terminal extein residues were amplified with the following pair of primers: prp8-if (5′-GGATCCTTCTGGGAGAAAGCGTGTCTC GC-3′) and prp8-ir (5′-GGATCCATGGTGATGGTGATG GTGAGATCCTCTTTCGAAACCACTGTTGTGCAG-3′). Both primers generated *Bam*HI sites at the 5′ ends, and a RGSHHHHHH-tag in case of the prp8-ir. The PCR amplification products containing the *prp8* intein coding sequence with flanking extein sequences (5aa and 4aa, respectively) of *P. chrysogenum*, *P. vulpinum*, and *P. expansum* fused to the RGS-His-tag were cloned into PCR cloning vector pDrive (Qiagen, Hilden, Germany). DNA sequences were confirmed by sequencing the PCR fragments (MWG-Biotech Customer Service, Ebersberg, Germany). Subsequent to sequencing, plasmids were digested with *Bam*HI and the 531-bp (*P. chrysogenum*), 546-bp (*P. expansum*), and 543-bp (*P. vulpinum*) intein-His-tag fragments were cloned into the single *Bam*HI sites of the GST-vector pGEX-4T-1 (Amersham Bioscience, Europe GmbH, Freiburg, Germany), respectively. The *Bam*HI fragments were either fused in frame with the *GST* gene of pGEX-4T-1 (pGPch-1, *P. chrysogenum*; pGPvu-1, *P. vulpinum*; pGPex-1, *P. expansum*) or fused in antisense orientation with the *GST* gene of pGEX-4T-1 (pGPch-2, *P. chrysogenum*; pGPvu-2, *P. vulpinum*; pGPex-2, *P. expansum*). A splicing incompetent variant of plasmid pGPex-1 was prepared by generating amino acid substitutions at the C-terminal splice site junction of the *P. expansum* intein. To replace the C-terminal Asn codon of the intein and the first residue of the C-terminal extein we used oligonucleotides prp8-if and prp8-ir-3 (5′-GGATCCATGGTGATGGTGATGGTGAGATCCTCTTTC GAAACC*AGCGTG*GTGCAG-3′) for PCR amplification of the *P. expansum prp8* intein. Oligonucleotide prp8-ir-3 leads

to an His/Ala substitution at the Asn/Ser wild-type intein/extein junction. To generate pGPex1-1, the amplified fragment was cloned in frame to the *GST* gene into the *Bam*HI site of pGEX-4T-1.

To introduce the intein of *P. expansum* into a split version of the GFP protein, we used plasmid pHGSapS (a kind gift from Henry Paulus, Boston Medical Research Institute) and constructed plasmid pHGPexS. Plasmid pHGSapS is identical to pHGSap described in (Gangopadhyay et al. 2003a), except for the substitution of the Ser codon TCG for the Cys codon TGT at positions 6195–6197 (Paulus, personal communication). Plasmid pHGSapS encodes a variant of GFP$_{UV}$ (Crameri et al. 1996) with a His-tag at the N-terminus, the substitutions E125V and I129S, and a nonapeptide insertion after I129S. The insertion is flanked by tandem *Sap*I sites. To construct pHGPexS, the *P. expansum* intein was amplified with oligonucleotides SaPex-1 (5′-GCTCTTC GTGCCTTGCCAAGGGAACCCGTCTA-3′) and SaPex-2 (5′-GCTCTTCGCGAGTTGTGCAGGACAAGGTAGTC-3′) generating *Sap*I recognition sites at both ends of the intein. The insertion of the PRP8 intein adjacent to GFP S129 was effected by the *Sap*I strategy previously described by Gangopadhyay et al. (Gangopadhyay et al. 2003b), using the coding region for the *P. expansum* intein flanked by *Sap*I sites to replace the nonapeptide insert in plasmid pHGSapS. The DNA sequence of all plasmids was verified by sequencing at MWG-Biotech Customer Service (Ebersberg, Germany).

Plasmid encoded proteins were overexpressed in *E. coli* strain BL21(DE3) Gold (Stratagene, La Jolla, CA, USA). Cultures were grown in liquid Luria Broth (LB) medium at 37°C with shaking to a culture density (A_{600}) of about 0.6 and incubated with 0.5 mM isopropyl-1-thio-β-D-galacto-pyranoside (IPTG) at 37°C for 4 h. Cells were harvested by centrifugation and lysed in SDS-containing gel-loading buffer in a boiling-water bath before electrophoresis in a 15% SDS-polyacrylamide gel.

The protein transfer to polyvinylidene difluoride Western blotting membrane (Biometra, Göttingen, Germany) was performed by using a semi-dry blotting system (Biometra, Göttingen, Germany). The detection was carried out with a (1:6,000) polyclonal anti-His-RGS antibody (Qiagen, Hilden, Germany) or with a (1:100,000) anti-GST antibody or with a (1:8,000) polyclonal anti-GFP living-colours peptide antibody (BD Bioscience, France) and with the chemiluminescence Western blotting kit (Roche, Germany), as described by the manufacturers.

GFP fluorescence measurements

Recombinant GFP fusion protein was isolated as described by Gangopadhyay et al. (2003a). Cultures (50 ml) of *E. coli* strain BL21(DE3) Gold (Stratagene, La Jolla, CA, USA) transformed with pHGSapS and pHGPexS were grown at 37°C in Luria broth supplemented with 100 μg/ml ampicillin. At a culture density (A_{600} nm) of 0.5, the cultures were induced with 0.5 mM IPTG and allowed to grow for another

4 h at 37°C. Cells were harvested by centrifugation at 6,500×*g* for 10 min, resuspended in 3 ml of buffer A (20 mM sodium phosphate, pH 7.5, 0.5 M NaCl) and disrupted by sonication. The pellet obtained by centrifugation at 16100×*g* for 20 min was resuspended in 2 ml of buffer B (buffer A supplemented with 8 M urea) to extract the inclusion bodies and centrifuged at 16,500×*g* for 10 min to remove the insoluble material. Renaturation of the purified inclusion bodies in 8 M urea was affected by dialysis at 4°C against 200 volumes of buffer C (20 mM sodium phosphate buffer, pH 7.0, supplemented with 0.5 M NaCl and 0.5 M arginine) with three buffer changes at 30-min intervals. GFP fluorescence was measured with a JASCO spectrofluorometer (Model FP-6500/6600, Jasco International Co., Japan) using an excitation at 395 nm by scanning the emission spectra between 450 and 600 nm. The band width for both excitation and emission spectra was 5 nm.

Results

The *prp8* gene of three Penicillium species contains an intein

To assess the occurrence of inteins in the *prp8* gene of species of the genus *Penicillium*, putative intein-containing fragments, according to the insertion site of previously identified *C. neoformans*, *A. nidulans*, *A. fumigatus*, and *H. capsulatum prp8* inteins (Butler et al. 2001; Liu and Yang 2004) were amplified from the genomic DNA of *P. chrysogenum*, *P. expansum*, *P. vulpinum*, and *P. thomii*. PCR amplifications were performed with oligonucleotides, allowing the detection of an intein sequence in an approximately 300-bp fragment of the *prp8* gene encoding the conserved 3.2 region of the PRP8 protein. The presence of an intervening sequence, an intein and/or intron sequence, can thus be revealed by the amplification of a longer PCR fragment. While the amplification of the *P. thomii* genomic DNA revealed a fragment of 300 bp, a fragment of approximately 800 bp was amplified from the genomic DNA of *P. chrysogenum*, *P. expansum*, and *P. vulpinum* (Fig. 1).

The *prp8* amplicons were cloned and their sequences were determined. The results confirmed an intron- and intein-less *prp8* fragment in *P. thomii* and the presence of an intervening sequence of 471, 486, and 483 bp in the *P. chrysogenum*, *P. expansum*, and *P. vulpinum prp8* genes, respectively. The deduced amino acid sequence revealed that they encode relatively small mini-inteins ranging in size from 157 to 162 amino acids. Similar to the PRP8 inteins of strains of the basidiomycete *C. neoformans*, but in contrast to PRP8 inteins of the filamentous ascomycetes *A. nidulans* and *A. fumigatus*, an endonuclease domain is missing in the *Penicillium* PRP8 inteins (Liu and Yang 2004).

The *Penicillium* inteins are located at exactly the same insertion site as inteins of *C. neoformans*, *A. fumigatus*, *A. nidulans*, and *H. capsulatum* (Butler et al. 2001; Liu and

a

b
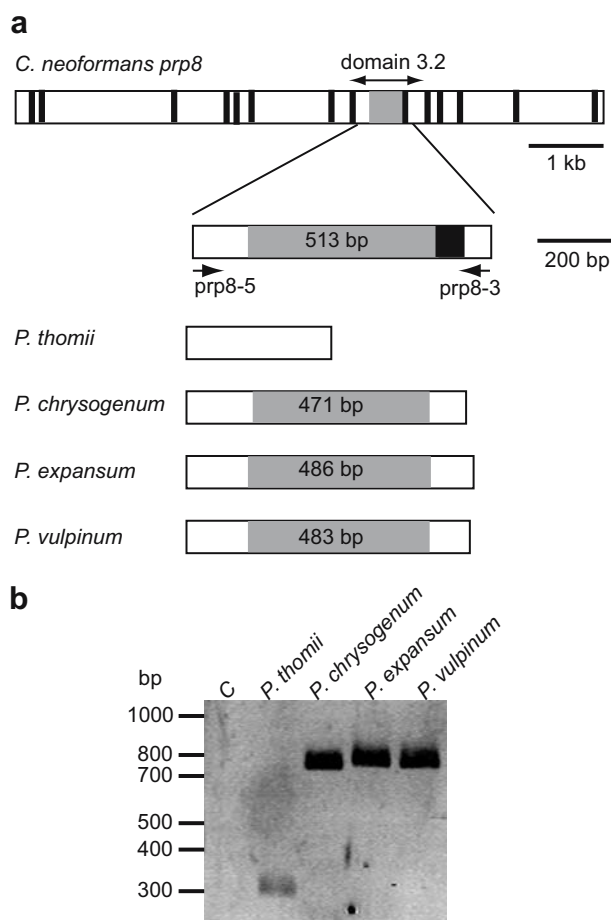
Fig. 1 PCR amplification of the *prp8* gene fragments from *P. thomii, P. chrysogenum*, *P. expansum* and *P. vulpinum*. **a** Schematic comparison of *prp8* fragments amplified from *P. chrysogenum*, *P. vulpinum*, *P. expansum*, and *P. thomii* with the *prp8* gene of *C. neoformans*. Inteins are indicated in *grey*, introns as *black bars*. **b** Gel electrophoresis of PCR amplicons produced with primer pair prp8-5 and prp8-3 from the genomic DNA of *P. thomii*, *P. chrysogenum*, *P. expansum*, and *P. vulpinum*. Lane C represents a negative control without DNA

Yang 2004). This insertion site is within an exceptionally well-conserved domain of PRP8 proteins, which is referred to as the 3.2 region. This region of the PRP8 protein is purportedly responsible for promoting RNA-mediated catalysis between the conserved 3′ splice site, the 5′ splice site, and the branchpoint of introns (Grainger and Beggs 2005). Apart from the conserved integration site, other sequence features indicate that the *Penicillium prp8* intervening sequences encode inteins: the inteins are in-frame with the PRP8 sequence and all amino acid residues providing nucleophilic groups in self-splicing reactions are present (Liu 2000; Noren et al. 2000). These include the first cysteine, the penultimate histidine, and the last asparagine residues of the inteins, as well as the first serine residue of the downstream extein. Moreover, conserved sequence blocks [A, B, F, and G (Pietrokovski 1994; Perler 2002)] were identified when the *Penicillium* inteins were aligned with known fungal PRP8 inteins (Fig. 2). Hereafter, the inteins identified in *P. chrysogenum*,

P. expansum, and *P. vulpinum* are referred to as Pch PRP8, Pex PRP8, and Pvu PRP8, respectively.

Analysis of protein splicing of Penicillium PRP8 inteins

To determine whether the *prp8* inteins of *P. chrysogenum*, *P. expansum*, and *P. vulpinum* are functional, we tested their protein-splicing activity in *E. coli* cells. Plasmid-borne fusion genes were constructed to produce proteins in which the *Penicillium* inteins (plus their 5-aa native N-extein sequences and 4-aa native C-extein sequence) were fused to an N-terminal GST protein and a C-terminal His-tag (Fig. 3). Similarly, fusion proteins consisting of an N-terminal maltose-binding protein and a C-terminal thioredoxin have been used in previous studies to analyze the activity of fungal PRP8 inteins (Liu and Yang 2004). A C-terminal His-tag as C-extein, together with an N-terminal maltose-binding protein, has been previously used to analyze splicing of a *Pyrococcus abyssi* intein (Mills et al. 2004).

In our experimental system, splicing in *E. coli* is expected to result in the conversion of the 49.9–51.0-kDa *G*ST-*I*ntein-*H*is-tag (GIH) precursor to the 32.2-kDa GH-spliced product and the 17.7–18.8-kDa excised intein (I). Cleavage at the N-terminal splice junction would be expected to result in the production of G (29.3 kDa) and IH (20.6 to 21.7 kDa).

SDS-PAGE analysis resulted in bands consistent with the formation of spliced products (GH) and excised inteins (I), which were identified by their predicted size. Precursor proteins, spliced products, and a splice intermediate consisting of the intein and the C-terminal His-tagged extein were further identified by using an anti-His-RGS antibody. The precursor (GIH) and the intein/C-extein splice intermediate (IH) of *E. coli* cells carrying pGPch-1 were only detected after a prolonged exposure, indicating the protein splicing was most efficient with the Pch PRP8 intein. The identity of His-tagged GST-splice products was also verified by Western blot analysis with an anti-GST antibody (Fig. 3). *Penicillium* inteins were also introduced in antisense orientation into the vector pGEX-4T-1. *E. coli* transformants carrying these constructs (pGPch-2, pGPex-2, or pGPvu-2) did not exhibit any protein-splicing activity. Only in the case of pGPvu-2 was the GST protein synthesized. Most probably because of instability of the GST fusion with the antisense constructs, *E. coli* transformants carrying pGPch-2 or pGPex-2 did not synthesize the GST protein at all.

Temperature did not influence the splicing efficiency of the *Penicillium* mini-inteins. Each of the *Penicillium* PRP8 inteins showed efficient protein splicing at room temperature as well as at 37 or 42°C with little or no precursor protein remaining (data not shown). Similarly, temperature-independent splicing has previously been reported for the *A. fumigatus*, *A. nidulans*, and *C. neoformans* PRP8 inteins. Only the *H. capsulatum* PRP8 intein has been shown to exhibit significantly less protein-splicing activity at 37°C (Liu and Yang 2004).

Fig. 2 ClustalX alignment of fungal PRP8 inteins. The 5-aa N-extein and 4-aa C-extein sequences are shown in *bold italics*. Conserved blocks of inteins are shaded in *grey* and *underlined*. Positions of the endonuclease domain (Endo) and the putative tongs domain of *A. fumigatus* (TSD) are indicated. The numbers of amino acids of these domains are given in parenthesis. The abbreviations and accession numbers for the intein sequences are: Cba, *Cryptococcus bacillisporus*, (Q6TEQ9); Cne_gr, *Cryptococcus neoformans var. grubii*, (Q6TER0); Cne_ne, *Cryptococcus neoformans var. neoformans*, (Q6TER1); Cne, *Cryptococcus neoformans* (Q8X1J5); Cla, *Cryptococcus laurentii* (Q58TW4); Pex, *Penicillium expansum* (AM042017); Pch, *Penicillium chrysogenum*, (AM042015), Pvu, *Penicillium vulpinum*, (AM042016); Ani, *Aspergillus nidulans* (AAX89368.1), Afu, *Aspergillus fumigatus* (AAV91021), Nfi, *Neosartorya fischeri* (AAT79351.1); Hca, *Histoplasma capsulatum* (Liu and Yang 2004). *Symbols* represent gaps introduced to optimise alignments; * mark identical amino acids, . mark similar amino acids

```
Cne_ne   FWEKACLQNGTRLLRADGSEVLVEDVQEGDQLLGPDGTSRTASKIVRGEERLYR---IKT
Cne      FWEKACLQNGTRLLRADGSEVLVEDVQEGDQLLGPDGTSRTASKIVRGEERLYR---IKT
Cne_gr   FWEKACLQNGTRLLRADGSEVLVEDVQEGDQLLGPDGTSRTASKIVRGEERLYR---IKT
Cba      FWEKACLQNGTRLLRADGSEILVEDVQEGDQLLGPDGTSRTASKIVRGEERLYR---IKA
Pvu      FWEKACLAKGTRLLRYDGTEINVEDVREGDQLLGPDGEPRRAFNIVSGIDRLYR---VKI
Pch      FWEKACLAKGTRLLRCDGTEINVEDVREGDQLLGPDGEPRRAFNIVNGIDRLYR---IKI
Pex      FWEKACLAKGTRLLRYDGTEINVEDVREGDQLLGPDGEPRRAFNIVSGIDRLYR---IKI
Nfi      FWERACLAKGTRLLRYDGSEIEVQDVKEGDLLLGPDGGPRRAFNIVNGEDRLYR---IKI
Afu      FWEKACLAKGTRLLRYDGSEIEVQDVKEGDLLLGPDGGPRRAFNIVNGKDRLYR---IKI
Ani      FWEKACLANGTQLLRYDGTKVNVEDVKEGDLLLGPDGGPRRAFNVVSGKDRLYR---IKI
Hca      FWEKACLAKGTQLLRYDGTKVGVENVREGDLLLGPDGEPRRAFNIVSGRDRLYR---ISI
Cla      FWEKSCHAEGTELIKFDRSVVKVEDVVEGDRLLGPDGSPRLVSNLVSGCDRLYRFEFSRS
         ***::*_ :**.*:: * : : *::* *** ****** .* . ::* * :****
                     A

Cne_ne   HEGLEDLVCTHNHILSMYKERSGSERAHSPSADLSLTDSHERVDVTVDDFVRLPQQEQQK
Cne      HEGLEDLVCTHNHILSMYKERSGSERAHSPSADLSLTDSHERVDVTVDDFVRLPQQEQQK
Cne_gr   HEGLEDLVCTHNHILSMYKERFGREGAHSPSAGTSLTESHERVDVTVDDFVRLPQQEQQK
Cba      DE-LEDLVCTHNHILSLYKKRSGSEQDPSPSTDLSSTDSYERVDVTVDDFVRLPQQEQQK
Pvu      GGEKEDLVVTPNHILVFYREG---------PSDGPENAERQTVEITAAEFATLSTEERSL
Pch      GGEKEDLVVTPNHILVLYRE------------DGSKNVEKQTVEITAAEFAALSTEERSL
Pex      AGEKEDLVVTPNHILVLYREEE--------ASDGPKNAERQTVEITAAEFAALSTEERGL
Nfi      DGSKEDLVVTPNHILVLHREKRA-----TTFESMPSTNAEEPGNGDDDELPEVSAAERYD
Afu      GGSKEDLVVTPNHILVLHR-----(TSD, 222 aa)-------GDDDELPQVSAEERYD
Ani      DGDKEDLVVTANHILVLHRAKAM-----NTSVCFDRSKEQQGGAGEQLDISEVSAAERYD
Hca      DADKEDLVVTPNHILVLHH---------------EKEN------QKRQSELSASATERYD
Cla      GGEVESLVVTDNHIMMLKRAN-----------VVLSGSGDAGQLEVLD--QISVREQFE
         *.** * ***: : :                             .    . *:
                    B

Cne_ne   YQLFRSTASVRH---------------ERPFTSKLDTTLLRINSIELEDEPTKWSGFVV
Cne      YQLFRSTASVRH---------------ERPSTSKLDTTLLRINSIELEDEPTKWSGFVV
Cne_gr   YKLFRSTDFVRR---------------EQPSASKLAT-LLHINSIELEEEPTKWSGFVV
Cba      YRLFRSTGFKRA---------------DQPSTSSLAT-LLHIMSIQLEEKPTKWSGFVV
Pvu      YSAFTSPAVEKG---------------AEGSAAQMHS--FKVEDISLESEKTEWAGFRV
Pch      YSAFTSPRAEKG---------------ADDS-AQTHS--FKIEQVSLESEKTEWAGFRV
Pex      HSAFTSSRVEKD---------------VENSAPQMHS--FKIEHINLEYEETEWAGFRV
Nfi      TVEMTAAEFAAL-(Endo, 354 aa)---RLIPQTHS--FMIKDISLEPETTEWAGFRV
Afu      TVEMTAAEFASL-(Endo, 454 aa)---RLIPQTHS--FMIKDISLEPEATEWAGFRV
Ani      TVEMTAAEFAAL-(Endo, 442 aa)---RLIPQSHS--FAIKDISLESEATEWAGFRV
Hca      TVEMTAADFAAL-(Endo, 387 aa)---SPVSQAHS--FTIKGIHLESEMTEWAGFRV
Cla      IVQKTAAEVAAM-(Endo, 365 aa)---RYLPQQHT--FRIKSVTLESKSTPWYGFAV
         :.                       .. : : :: : ** : * * ** *
                                                            F

Cne_ne   DKDSLYLRHDYLVLHNSGFE
Cne      DKDSLYLRHDYLVLHNSGFE
Cne_gr   DKDSLYLRYDYLVLHNSGFE
Cba      DKDSLYLRHDYLVLHNSGFE
Pvu      DKDQLYLRHDYLVLHNSGFE
Pch      DKDQLYLRHDYLVLHNSGFE
Pex      DKDQLYLRHDYLVLHNSGFE
Nfi      DKDQLYLRHDYLVLHNSGFE
Afu      DKDQLYLRHDYLVLHNSGFE
Ani      DKDQLYLRHDYVVLHNSGFE
Hca      DKDQLYLRHDFLVLHNSGFE
Cla      DSDRQYLRHDWMVLHNSGFE
         *.*   ***:*::*******
                  G
```

Amino acid substitution at the intein/C-extein junction prevent splicing of the Pex PRP8 intein

To confirm that the splicing mechanism of the *Penicillium* PRP8 inteins involves the four steps of the splicing procedure identified for the majority of inteins, the activity of a mutated Pex PRP8 intein (encoded by pGPex1-1) was analyzed. Substitution of the conserved intein terminal Asn by His and the C-terminal flanking residue Ser by Ala completely prevented splicing activity (Fig. 4). SDS-PAGE and Western blot analysis using the anti-His-RGS antibody detected a band consistent with the production of massive amounts of GIH precursor, however, no cleavage or splicing products. Additional bands referring to cleavage or splicing products remained also undetectable with use of the anti-GST antibody (data not shown). Under the same conditions, fusion of wt Pex PRP8 intein to the N-terminal GST and the C-terminal His-tag resulted in the formation of the splicing and cleavage products GH, IH, and I. Thus, our result having no cleavage or splicing products after mutation of the terminal Asn and the C-extein Ser +1 suggests that Pex PRP8 intein protein splicing requires the conserved residues.

The Pex PRP8 intein can excise itself from a foreign host protein

To investigate whether the protein-splicing mechanism of the Pex PRP8 intein is only achieved by the intein sequence alone and does not require host protein residues immediately adjacent to the intein, the splicing activity of the Pex PRP8 intein was examined by placing it into the foreign
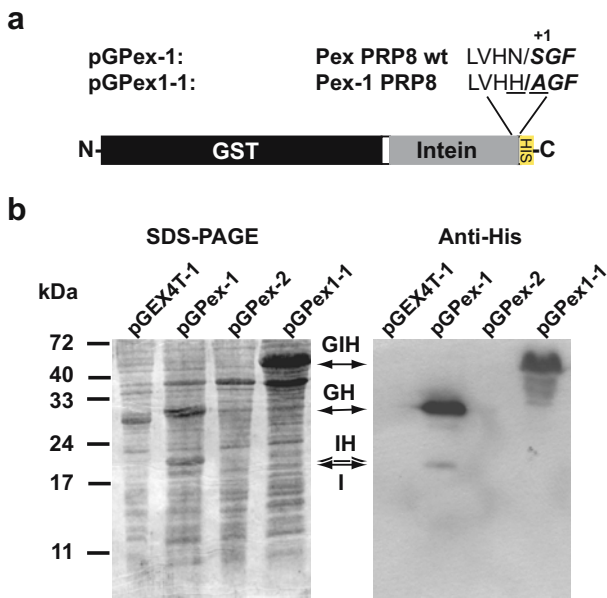
Fig. 3 Protein splicing in *E. coli*. **a** Schematic illustration of vectors used in this analysis. **b** Schematic diagram depicting fusion protein of glutathion S-transferase (*GST*), short extein sequences (*white*), intein and His-tag (*HIS*). Predicted sizes of fusion protein precursors (*GIH*), the splice products (*GH*), inteins (*I*), and the cleavage products (*IH*) and (*G*) are given for the inteins of *P. chrysogenum* (*Pch*), *P. expansum* (*Pex*), and *P. vulpinum* (*Pvu*). **c** SDS-PAGE and Western blot analysis of protein splicing. Recombinant *E. coli* strains contain as a control plasmid pGEX-4T-1, or they contain plasmid pGPch-1, pGPch-2, pGPex-1, pGPex-2, pGPvu-1, or pGPvu-2. After induction with IPTG, total cellular proteins of *E. coli* were resolved by SDS-PAGE and visualized by Coomassie Blue or by Western blotting using an anti-His-RGS antibody or anti-*GST* antibody. Prolonged exposure of lane pGPch-1 after incubation with the anti-His-RGS antibody is shown rightmost. Positions of precursors, spliced products, inteins, and splice intermediates were indicated
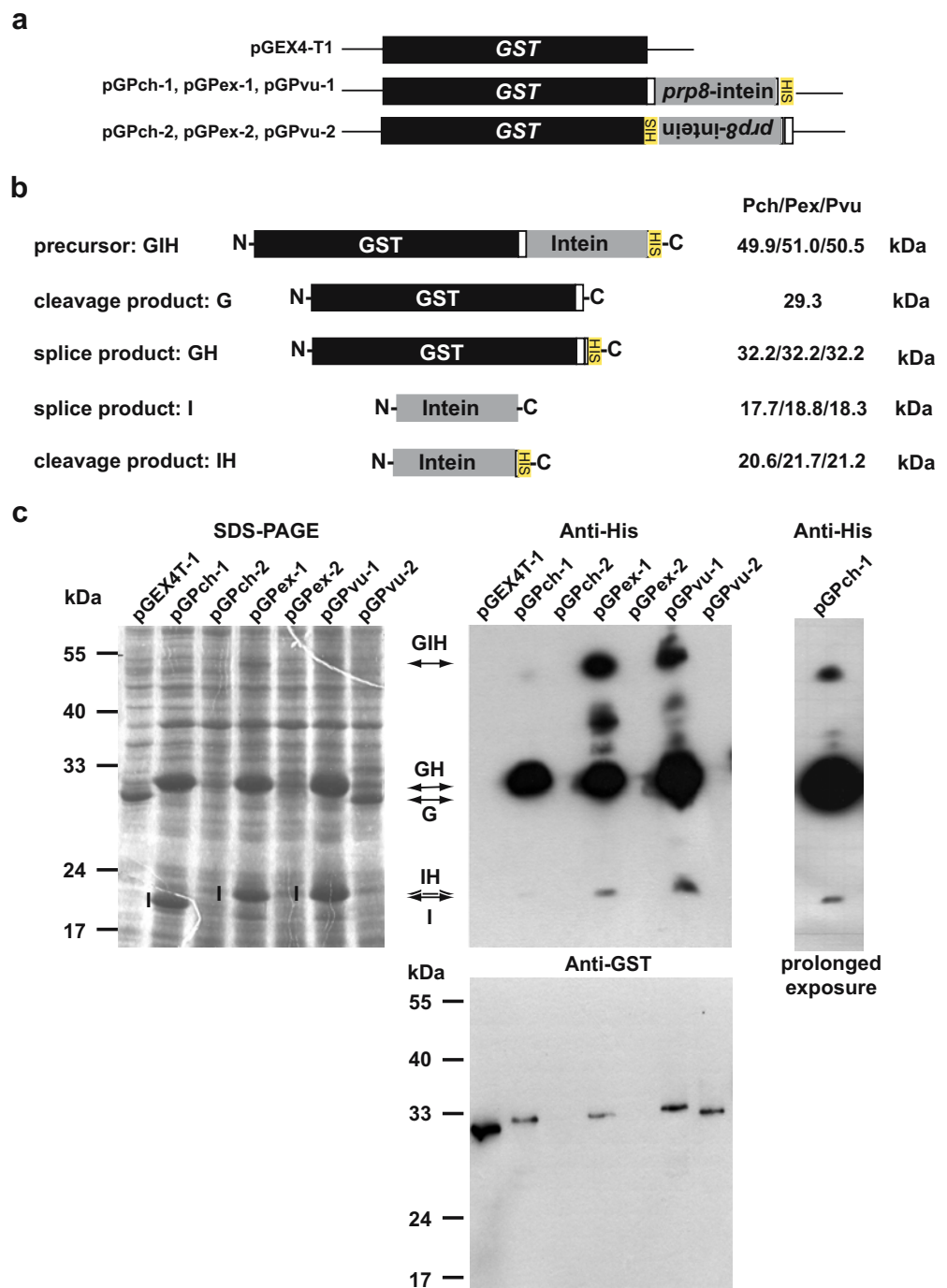
GFP protein. For this purpose, we cloned the Pex PRP8 intein without flanking PRP8 extein sequences into the *Sap*I site of the *gfp* vector pHGSapS and generated vector pHGPexS. Plasmid pHGSapS was used in this experiment because it encodes a variant of GFP$_{UV}$ (Crameri et al. 1996) with an Ile-to-Ser substitution at amino acid position 129. This choice assured that after cloning of the Pex PRP8 into the *Sap*I site of pHGSapS, the first C-terminal extein residue was a Ser residue, which was demonstrated to be important for splicing activity of Pex PRP8 (Fig. 5a). Previously, it was demonstrated that GFP fails to form its chromophore in vivo when interrupted by an intein inserted adjacent to residue 129, but that subsequent protein splicing allows fluorescence to develop (Ozawa et al. 2000; Gangopadhyay et al. 2003a). Thus, polypeptides inserted at this position cause the protein to form inclusion bodies, and only when the inserted polypeptide is an intein the renatured fusion protein can undergo protein splicing and chromophore formation. After induction with IPTG at 37°C, extracts of *E. coli* cells transformed either with pHGSapS or with pHGPexS were analyzed by SDS-PAGE and Western blot analysis (Fig. 5b). As expected,

Fig. 4 Mutagenesis study on conserved residues of the Pex PRP8 intein. **a** Schematic diagram depicting fusion protein of glutathion S-transferase (GST), short extein sequences (*white*), intein and His-tag (*HIS*). Conserved residues of the intein C-terminus and of +1 position of the C-terminal extein (*bold italics*) are shown for the Pex PRP8 wild type (*wt*). Amino acid substitutions of conserved residues in the mutated Pex-1 PRP8 intein/C-terminal extein are underlined. **b** SDS-PAGE and Western blot analysis of protein splicing. Recombinant *E. coli* strains contain as a control plasmid pGEX-4T-1, or they carry plasmid pGPex-1, pGPex-2, or pGPex1-1. After induction with IPTG, total cellular proteins of *E. coli* were resolved by SDS-PAGE and visualized by Coomassie Blue or by Western blotting using an anti-His-RGS antibody. Positions of precursors (*GIH*), spliced products (*GH*), inteins (*I*), and splice intermediates (*IH*) are indicated

plasmid pHGSapS led to the synthesis of a 31.2-kDa GFP, whereas, plasmid pHGPexS led to the synthesis of a 49.1-kDa GFP-intein precursor and small amounts of the spliced 30.3-kDa GFP protein. The identity of the precursor and the spliced GFP was verified by Western blot analysis with an anti-GFP antibody. In addition, we detected a band of about 13 kDa in the Western blot, which corresponded to the size of the C-terminal GFP extein. This GFP derivative might result from a C-terminal cleavage event of the precursor. A faint band of approximately 36 kDa might represent the N-terminal GFP-intein moiety of this cleavage event. When inclusion body fractions from *E. coli* transformed with plasmids pHGPexS were extracted with 8 M urea and renatured by dialysis GFP, fluorescence was obtained (Fig. 5c). In contrast, no fluorescence was observed with renatured inclusion bodies from plasmid pHGSapS transformants. Previously, it was already demonstrated by Gangopadhyay et al. (2003a) that the small 9-aa peptide insertion at residue 129 interferes with chromophore formation. The fluorescence observed upon renaturing the inclusion bodies encoded by plasmid pHGPexS was, therefore, due to prior excision of the Pex intein by protein splicing.

Discussion

We have identified PRP8 mini-inteins in *P. chrysogenum*, *P. expansum*, and *P. vulpinum*. To our knowledge, the *Penicillium* PRP8 inteins are among the smallest nuclear-encoded inteins identified so far. Only a 155-amino-acid PRP8 mini-intein of the filamentous ascomycete *Neosartorya glabra* (AAX39414.1) identified in a BLAST search was smaller than the 157-amino-acid *P. chrysogenum* intein. When compared with each other, all three *Penicillium* inteins display a high degree of sequence conservation. Amino acid identity of the splicing domain among species of the genus *Penicillium* ranges from 84.6% (between *P. expansum* and *P. chrysogenum*) to 86.4% (between *P. expansum* and *P. vulpinum*). Pairwise comparison of PRP8 inteins from different fungal organisms after excluding the homing endonuclease domain and the putative tongs domain (Werner et al. 2002; Liu and Yang 2004) revealed a high degree of amino acid conservation within the protein-splicing domain of PRP8 inteins. Sequence identity of the splicing domain is even over 60% between *Penicillium* species and other filamentous ascomycetes, and ~48% among species of the genus *Penicillium* and different *Cryptococcus* strains and species in the basidiomycete phylum. Because of the high degree of sequence conservation of fungal PRP8 inteins and their identical insertion site, a common origin of the PRP8 inteins has been suggested previously (Liu and Yang 2004; Butler et al. 2005).

The *Penicillium* inteins were shown to be active in protein splicing in *E. coli*. Mutagenesis of a conserved intein residue and the first C-terminal extein residue revealed that at least for the splicing of the *P. expansum* PRP8 intein, these amino acids are essential. For inteins derived from other organisms, it has been shown that substitution of amino acids located at three conserved positions at the N- and C-terminus of the intein and the amino acid directly following the intein C-terminus blocks protein splicing but still produces some side products. Substituting the C-terminal Asn into Ala, into Gln, or into Asp has previously been shown to block C-terminal cleavage of different inteins while still allowing N-terminal junction cleavage and branched intermediate formation (Xu and Perler 1996; Shingledecker et al. 2000; Southworth et al. 2000). However, in addition to the N-terminal Asn residue of the intein, we substituted the +1 Ser residue of the C-terminal *P. expansum* extein. According to the standard splicing pathway, the C-terminal nucleophile (Ser +1) is predicted to cleave the N-terminal splice junction. Mutation of Ser +1 should, thus, block the N-terminal splice junction cleavage observed in our analysis (Noren et al. 2000). Thus, it seems that the Pex intein follows the four-step splicing mechanism described for the majority of inteins (Liu 2000; Noren et al. 2000). The high degree of sequence conservation makes it very likely that the protein-splicing mechanism of Pch PRP8 and Pvu PRP8 is very similar to Pex PRP8 splicing.

Insertion of the Pex PRP8 intein into the foreign GFP protein revealed that, similar to inteins from other

organisms, the *P. expansum* PRP8 mini-intein with the first downstream extein residue contains sufficient information for splicing in foreign proteins (Davis et al. 1992; Cooper et al. 1993; Xu et al. 1993). Western blot analysis revealed that several side products occurred during protein splicing of the GFP-intein fusion protein. These products might be abortive single-splice junction cleavage products that have often been observed when inteins have been inserted into foreign proteins (Davis et al. 1992; Cooper et al. 1993; Xu et al. 1993; Telenti et al. 1997; Chong et al. 1998). It has been suggested that they might result from suboptimal proximal extein sequences chemically or sterically interfering with packing at the active site (Noren et al. 2000). However, although surrounded by an altered proximal extein sequence, the GFP-Pex PRP8 intein fusion exhibited a clearly visible splicing activity. Upon solubilisation in 8 M urea followed by renaturation, the heterologously expressed GFP-Pex intein fusion protein was able to undergo protein splicing to yield GFP capable of undergoing the rearrangements that lead to the formation of the fluorescent chromophore (Fig. 5c).

Previously, it has been suggested that fungal PRP8 inteins are not tolerated in mature PRP8 proteins, making them attractive as drug targets (Liu and Yang 2004). Our finding of protein splicing of the Pex PRP8 intein inside a modified GFP may be used to design an in vitro screening system for protein-splicing inhibitors of fungal PRP8 inteins (Gangopadhyay et al. 2003a; Paulus 2003). Because of their compactness and high splicing activity inside foreign proteins, *Penicillium* PRP8 mini-inteins provide a paradigm for a minimal protein-splicing element in filamentous fungi. Moreover, they may be used for the development of new intein-mediated protein-engineering applications in fungal systems, such as protein purification, addition of fluorescent biosensors, and expression of cytotoxic proteins (Noren et al. 2000; Perler 2005).

Temperature-sensitive splicing variants may be used for the generation of temperature-sensitive alleles, which were recently demonstrated to represent useful tools for the investigation of gene function (Zeidler et al. 2004).

References

Belfort M, Roberts RJ (1997) Homing endonucleases: keeping the house in order. Nucleic Acids Res 25:3379–3388

Butler MI, Goodwin TJ, Poulter RT (2001) A nuclear-encoded intein in the fungal pathogen *Cryptococcus neoformans*. Yeast 18:1365–1370

Butler MI, Goodwin TJ, Poulter RT (2005) Two new fungal inteins. Yeast 22:493–501

Chong S, Williams KS, Wotkowicz C, Xu MQ (1998) Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. J Biol Chem 273:10567–10577

Cooper AA, Chen YJ, Lindorfer MA, Stevens TH (1993) Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. EMBO J 12:2575–2583

Crameri A, Whitehorn EA, Tate E, Stemmer WP (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol 14:315–319

Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedgwick SG (1992) Protein splicing in the maturation of *M. tuberculosis* recA protein: a mechanism for tolerating a novel class of intervening sequence. Cell 71:201–210

Elander RP (2003) Industrial production of beta-lactam antibiotics. Appl Microbiol Biotechnol 61:385–392

Gangopadhyay JP, Jiang SQ, Paulus H (2003a) An in vitro screening system for protein splicing inhibitors based on green fluorescent protein as an indicator. Anal Chem 75:2456–2462

Gangopadhyay JP, Jiang SQ, van Berkel P, Paulus H (2003b) In vitro splicing of erythropoietin by the *Mycobacterium tuberculosis* RecA intein without substituting amino acids at the splice junctions. Biochim Biophys Acta 1619:193–200

Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E (2002) Inteins: structure, function, and evolution. Annu Rev Microbiol 56:263–287

Grainger RJ, Beggs JD (2005) Prp8 protein: at the heart of the spliceosome. RNA 11:533–557

Greener A (1990) New competent cells for highest transformation efficiencies. Strategies 3:5–6

Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y (1990) Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. J Biol Chem 265:6726–6733

Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science 250:651–657

Liu XQ (2000) Protein-splicing intein: Genetic mobility, origin, and evolution. Ann Rev Genet 34:61–76

Liu XQ, Yang J (2004) Prp8 intein in fungal pathogens: target for potential antifungal drugs. FEBS Lett 572:46–50

Majerus P, Kapp P (2002) Assessment of dietary intake of patulin by the population of EU member states. http://europa.eu.int/comm/food/fs/scoop/3.2.8_en.pdf:1-138

Mills KV, Manning JS, Garcia AM, Wuerdeman LA (2004) Protein splicing of a *Pyrococcus abyssi* intein with a C-terminal glutamine. J Biol Chem 279:20685–20691

Noren CJ, Wang J, Perler FB (2000) Dissecting the chemistry of protein splicing and its applications. Angew Chem Angew Chem Int Ed Engl Suppl 39:450–466

Ozawa T, Nogami S, Sato M, Ohya Y, Umezawa Y (2000) A fluorescent indicator for detecting protein–protein interactions in vivo based on protein splicing. Anal Chem 72:5151–5157

Paulus H (2000) Protein splicing and related forms of protein autoprocessing. Annu Rev Biochem 69:447–496

Paulus H (2003) Inteins as targets for potential antimycobacterial drugs. Front Biosci 8:1157–1165

Perler FB (2002) InBase: the intein database. Nucleic Acids Res 30:383–384

Perler FB (2005) Protein splicing mechanisms and applications. IUBMB Life 57:469–476

Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M (1994) Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. Nucleic Acids Res 22:1125–1127

Pietrokovski S (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. Protein Sci 3:2340–2350

Pietrokovski S (2001) Intein spread extinction in evolution. Trends Genet 17:465–472

Pitt JI (1979) The genus *Penicillium* and its teleomorphic states *Eupenicillium* and *Talaromyces*. Academic, London

Pöggeler S (2005) Fungal intervening sequences. In: D Arora, Khachatourians GG (ed) Applied mycology and biotechnology. Elsevier Sci, Oxford

Pöggeler S, Nowrousian M, Jacobsen S, Kück U (1997) An efficient procedure to isolate fungal genes from an indexed cosmid library. J Microbiol Methods 29:49–61

Sambrook J, Fritsch EF, Maniatis T (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Shingledecker K, Jiang S, Paulus H (2000) Reactivity of the cysteine residues in the protein splicing active center of the *Mycobacterium tuberculosis* RecA intein. Arch Biochem Biophys 375:138–144

Southworth MW, Benner J, Perler FB (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. EMBO J 19:5019–5026

Telenti A, Southworth M, Alcaide F, Daugelat S, Jacobs WRJ, Perler FB (1997) The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. J Bacteriol 179:6378–6382

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Walz M, Kück U (1995) Transformation of *Sordaria macrospora* to hygromycin B resistance: characterization of transformants by electrophoretic karyotyping and tetrad analysis. Curr Genet 29:88–95

Werner E, Wende W, Pingoud A, Heinemann U (2002) High resolution crystal structure of domain I of the *Saccharomyces cerevisiae* homing endonuclease PI-SceI. Nucleic Acids Res 15:3962–3971

Xu MQ, Perler FB (1996) The mechanism of protein splicing and its modulation by mutation. EMBO J 15:5146–5153

Xu MQ, Southworth MW, Mersha FB, Hornstra LJ, Perler FB (1993) In vitro protein splicing of purified precursor and the identification of a branched intermediate. Cell 75:1371–1377

Zeidler MP, Tan C, Bellaiche Y, Cherry S, Hader S, Gayko U, Perrimon N (2004) Temperature-sensitive control of protein activity by conditionally splicing inteins. Nat Biotechnol 22:871–876