Control of protein splicing by intein fragment reassembly

Maurice W. Southworth, Eric Adam, Daniel Panne¹, Robyn Byer, Roger Kautz² and Francine B. Perler³

New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

¹Present address: Biozentrum, Abteilung Mikrobiologie, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

²Present address: Barnett Institute, Northeastern University, Boston, MA 02115, USA

³Corresponding author
e-mail: perler@neb.com

Inteins are protein splicing elements that mediate their excision from precursor proteins and the joining of the flanking protein sequences (exteins). In this study, protein splicing was controlled by splitting precursor proteins within the Psp Pol-1 intein and expressing the resultant fragments in separate hosts. Reconstitution of an active intein was achieved by *in vitro* assembly of precursor fragments. Both splicing and intein endonuclease activity were restored. Complementary fragments from two of the three fragmentation positions tested were able to splice *in vitro*. Fragments resulting in redundant overlaps of intein sequences or containing affinity tags at the fragmentation sites were able to splice. Fragment pairs resulting in a gap in the intein sequence failed to splice or cleave. However, similar deletions in unfragmented precursors also failed to splice or cleave. Single splice junction cleavage was not observed with single fragments. *In vitro* splicing of intein fragments under native conditions was achieved using mini exteins. *Trans*-splicing allows differential modification of defined regions of a protein prior to extein ligation, generating partially labeled proteins for NMR analysis or enabling the study of the effects of any type of protein modification on a limited region of a protein.

Keywords: intein/protein expression/reconstitution/split proteins/thermophile/urea

Introduction

Protein splicing is a post-translational process that results in excision of an intein (protein splicing element) from a precursor protein and the ligation of the flanking protein sequences (exteins) to yield two mature proteins, the intein and the ligated exteins (Perler *et al.*, 1994). The native peptide bond formed between the exteins (Cooper *et al.*, 1993) distinguishes protein splicing from other forms of autoprocesing (Perler *et al.*, 1997b). The self-catalytic reaction requires four nucleophilic displacements mediated by three conserved splice junction residues: (i) a Ser or Cys at the intein N-terminus; (ii) an Asn at the intein C-

terminus; and (iii) a Ser, Thr or Cys at the beginning of the C-extein (Xu *et al.*, 1993, 1994; Shao *et al.*, 1995, 1996; Chong *et al.*, 1996; Xu and Perler, 1996). Genetic, biochemical and structural studies have shown that formation of the splicing active site requires proper folding of the intein to bring together the two splice junctions that can be >500 amino acids apart, plus other intein residues that may assist in the nucleophilic displacements, such as the conserved His in intein blocks B and G (Petrokovski, 1994, 1996; Duan *et al.*, 1997; Hall *et al.*, 1997; Kawasaki *et al.*, 1997; Perler *et al.*, 1997a,b).

Many inteins are bifunctional proteins, having both splicing and homing endonuclease activity (Bremer *et al.*, 1992; Perler *et al.*, 1992; Mueller *et al.*, 1994). The mature Psp Pol-1 intein is also a homing endonuclease, PI-PspI, that specifically cleaves the intein insertion site in DNA polymerase genes that lack the intein (F. Perler, unpublished data). This type of homing endonuclease activity is thought to initiate intein gene mobility into inteinless extein alleles (Mueller *et al.*, 1994; Perler *et al.*, 1997a).

In an attempt to control splicing and allow differential labeling or modification of portions of a protein, we split several precursors within the Psp Pol-1 intein and examined whether splicing could be reconstituted *in vitro* from the separately purified parts (Figure 1). Limited proteolysis experiments have proven that folded proteins can remain active despite the presence of breaks in the peptide backbone (Anfinsen and Scheraga, 1975). Previous studies have also indicated that under certain conditions, protein fragments are able to find their complementary partners and fold properly to generate an active enzyme (Kato and Anfinsen, 1969; Matsuyama *et al.*, 1990; Burbaum and Schimmel, 1991; Sancho and Fersht, 1992; Kanaya and Kanaya, 1995; Tasayco and Chao, 1995; Gross *et al.*, 1996). These and other studies also demonstrated that the conformation of protein fragments is often disordered, and hydrophobic regions that are normally buried in the intact protein may be exposed, leading to aggregation, insolubility or *in vivo* proteolysis. Most *in vitro* assembly protocols include a denaturation step prior to or during fragment association (Kato and Anfinsen, 1969; Anfinsen and Scheraga, 1975; Matsuyama *et al.*, 1990; Burbaum and Schimmel, 1991; Sancho and Fersht, 1992; Kanaya and Kanaya, 1995; Tasayco and Chao, 1995). Finally, *in vivo* reconstitution is often more efficient than *in vitro* reconstitution (Gross *et al.*, 1996). *In vivo* reassembly can be aided by reassociation before the co-expressed fragments misfold and/or by the assistance of the powerful protein folding machinery present in the cell. However, *in vivo* assembly does not allow differential labeling or modification of portions of a protein, nor does it necessarily block protein splicing *in vivo*.

In this study, we examined the ability of an enzyme from an extreme thermophile (*Pyrococcus* species, isolate

Fig. 1. Scheme for reassembly of split inteins. The MIP protein splicing precursor (Xu *et al.*, 1993) was split at various locations within the intein (I) to yield N-terminal fragments (MI_N) and C-terminal fragments ($I_C P$). However, precursors with other N-exteins or C-exteins can likewise be split within the intein. Purification tags can also be added to the intein split sites. After purification, the two fragments are mixed in buffers containing various amounts of urea (0–8.0 M) and allowed to reassemble. The reconstituted intein then directs the splicing reaction, resulting in joining of the extein fragments with a native peptide bond. The reconstituted intein fragments also display PI-PspI endonuclease activity.

GB-D) to assemble at temperatures of up to 100°C below its normal synthesis and folding temperatures. Assembly into an active enzyme was monitored by assaying protein splicing or endonuclease activity. The ability to reassemble intein fragments into an active enzyme converts any intein into a controllable protein splicing element. Moreover, it provides a method for the specific labeling or modification (e.g. phosphorylation, glycosylation, acetylation) of a protein fragment prior to assembly, allowing the analysis of the effects of the specific modification.

Results

Construction and expression of split precursor proteins

Intein fragmentation studies were performed with the previously characterized chimeric protein MIP which is a three part fusion of the *Escherichia coli* maltose-binding protein (M or MBP, the N-extein), the Psp Pol-1 intein (I) and the Δ Sal fragment of *Dirofilaria immitis* paramyosin (P, the C-extein) (Xu *et al.*, 1993). Splicing of MIP is optimal at pH 6–7 with a half-time of 20–30 min *in vitro* and is inhibited at low temperatures (4–16°C) or pH values above pH 9 (Xu *et al.*, 1993, 1994; Xu and Perler, 1996).

Since there are no precise rules for choosing split sites (the position at which the protein is split into fragments) (Matsuyama *et al.*, 1990; Burbaum and Schimmel, 1991; Gross *et al.*, 1996), three positions within the Psp Pol-1 intein were tested. No structural information was available for any intein or homing endonuclease at the inception of this project, although sequences of several alleles of the Psp pol-1 intein were available for comparison (Perler *et al.*, 1997a). We hypothesized that non-conserved, unstructured surface locations might be less essential to

the intein and, therefore, breakage of the peptide backbone in these regions might be less detrimental. Therefore, the three split sites were chosen in highly variable regions of the Psp Pol-1 intein that were predicted by computer modeling (Rost *et al.*, 1994) to be in unstructured loops with potential surface locations. Precursor proteins were split following intein residues Glu108 (MI_{N1} and $I_{C1}P$), Leu249 (MI_{N2} and $I_{C2}P$) and Arg440 (MI_{N3} and $I_{C3}P$) at split sites 1, 2 and 3, respectively (Figure 2). Leu249 precedes a naturally occurring Met at position 250, and Arg440 is near a protease-sensitive site at Lys442 (J. Benner and T. Davis, personal communication).

N-terminal precursor fragments (MI_{N1} , MI_{N2} and MI_{N3}) were synthesized as soluble proteins (10–40 mg/l) while C-terminal precursor fragments ($I_{C2}P$, $I_{C2}\Delta$ and $I_{C3}P$) were synthesized as insoluble proteins (30–40 mg/l). $I_{C1}P$ was very sensitive to *in vivo* proteolysis, but small amounts of full-length $I_{C1}P$ protein (0.5 mg/l) could be isolated under certain induction conditions. To eliminate the possibility that the insolubility of $I_C P$ fragments was due to the paramyosin domain, the paramyosin extein was replaced in $I_{C3}P$ by *E. coli* thioredoxin or the chitin-binding domain from the *Bacillus circulans* chitinase; both new C-terminal fragments were also insoluble.

Time course of splicing *in trans* confirms the protein splicing pathway

Splicing of precursor fragments *in trans* was successful when MIP was split at sites 2 and 3, after Leu249 or Arg440, respectively, but not at site 1, after Glu108 (Figures 2–5 and Table I). This percentage of successful fragment reassembly is similar to that reported in other systems (Matsuyama *et al.*, 1990; Burbaum and Schimmel, 1991). Products that would be unique to cleavage reactions were not observed. For example, I_N is a potential product of both splicing and cleavage while free M can only result from cleavage of MI_N . Similar results were also observed with intact MIP, where cleavage only occurred *in vivo* or after mutagenesis and not *in vitro*.

Splicing *in trans* confirmed the order of splice junction cleavage in the protein splicing pathway since such cleavage releases identifiable intein fragments in SDS-PAGE. The protein splicing pathway begins with ester formation followed by cleavage at the N-terminal splice site resulting in formation of a slowly migrating branched intermediate (MIP^*) containing M connected to the side chain of S538 in IP (Xu *et al.*, 1993, 1994; Shao *et al.*, 1995, 1996; Xu and Perler, 1996). In *trans*-splicing reactions, N-terminal splice site cleavage would be detected in SDS-PAGE by the appearance of I_N and the branched intermediate, $MI_C P^*$. The next step in the pathway is resolution of the branched intermediate by Asn cyclization which would lead to cleavage at the C-terminal splice site. In *trans*-splicing, this would result in the release of I_C and the ligated exteins (MP) from $MI_C P^*$. The MI_{N3} plus $I_{C3}P$ time course shown in the left panel of Figure 3 illustrates this splicing pathway. The first products observed are the branched intermediate ($MI_{C3} P^*$) and cleaved I_{N3} . Small amounts of both were present after the overnight incubation in 3.6 M urea (0 min splicing reaction sample). As the splicing reaction continued, more branched intermediate and I_{N3} were formed, followed by the appearance of the spliced product (MP) and I_{C3} . By 90 min, most of $MI_{C3} P^*$

Fig. 2. Map of MIP fragments. Restriction enzyme sites within the MIP gene used to construct subclones are shown across the top of the MIP precursor, and residues surrounding the Psp Pol-1 intein splice junctions (/) are shown below the MIP precursor. The eight conserved intein motifs are depicted, including splicing motifs (blocks A, B, F and G) and endonuclease motifs (blocks C, D, E and H). MIP was split at three sites after intein amino acids E108, L249 and R440 to generate three sets of complementary fragments. Fragment names are listed to the left of each fragment pair and include an N or C subscript indicating an N- or C-terminal intein sequence, respectively, and a subscript split site number (1–3). The terminal intein residue at each split site is shown above the fragment, and the number of intein amino acids in each fragment is listed within the white intein box. *In vitro* splicing activity of complementary fragments is shown to the right of each fragment pair. An MBP affinity tag was fused to the N-terminus of IC₃P to generate MI_{C3}P. Splicing activity of MI_{C3}P was assayed with its complementary fragment, MI_{N3}. Abbreviations: M or MBP, maltose-binding protein; P, paramyosin; (/) splice junction; white box, intein or intein fragment; shaded box, MBP or paramyosin exteins.

Fig. 3. Time course of splicing with MIP split at site 3. Splicing of the MI_{N3} plus IC₃P complementary pair of fragments proceeded as predicted by the protein splicing mechanism, including formation of a slowly migrating intermediate (MI_{C3}P*, left panel or MMI_{C3}P*, right panel), irrespective of the presence of an MBP affinity tag at the split site in MI_{C3}P. Left panel: MI_{N3} was mixed with IC₃P in 3.6 M urea buffer at 4°C overnight (protocol 2) and then diluted 10-fold into splicing buffer followed by incubation at 37°C for 0–90 min. Right panel: MI_{N3} was mixed with MI_{C3}P containing an MBP tag at the split site, treated as above and incubated for 0–120 min at 37°C. Lane 120(–) was not pre-incubated, but instead the fragments in amylose column elution buffer were diluted directly into splicing buffer and incubated at 37°C immediately after fragment mixing. Abbreviations: MI_{N3} and IC₃P or MI_{C3}P, substrate fragments; MP, spliced product; I_{N3}, IC₃ or MI_{C3}, intein fragment products. The SDS-PAGE gels were stained with Coomassie blue.

and the MI_{N3} substrate had disappeared, leaving some unreacted IC₃P substrate which was in molar excess.

Conditions for functional assembly of split precursors

Most *in vitro* protocols for assembly of split proteins involve denaturation followed by renaturation (Kato and

Fig. 4. Effect of urea in the splicing buffer. MI_{N2} and IC₂PA complementary fragments were pre-incubated in 6.0 M urea buffer at 4°C and then diluted 10-fold into splicing buffer containing 6 M urea (protocol 3, left panel) or 0 M urea (protocol 1, right panel) followed by incubation at 37°C for 0–120 min. The SDS-PAGE gel was stained with Coomassie blue. The presence of urea in the splicing buffer had no significant effect on the production of spliced MPΔ.

Anfinsen, 1969; Matsuyama *et al.*, 1990; Burbaum and Schimmel, 1991; Sancho and Fersht, 1992; Kanaya and Kanaya, 1995; Tasayco and Chao, 1995; Gross *et al.*, 1996). Several protocols for intein fragment assembly were examined (Table I). Parameters that were varied in this study included urea concentration (0–8.0 M), temperature (4 or 37°C) and length (0–20 h) of the pre-incubation step, and urea concentration (0–8.0 M) of the splicing step. The splicing step consisted of renaturation by rapidly diluting the pre-incubation mixtures ≥10-fold in splicing buffer (without urea, except in protocol 3) and incubating at 37°C to stimulate splicing. Splicing was monitored by observing the disappearance of MI_N and IC_P substrates and the appearance of MP, I_N and I_C products.

However, splicing efficiency was calculated only with respect to the synthesis of MP, since this is the desired spliced product. Significant amounts of either or both substrate fragments often remained at the end of the splicing reaction, partially because an equal amount of each substrate does not represent equal moles of fragments (due to differences in molecular weight). Misfolding and/or aggregation of substrate may also contribute to the

failure of splicing reactions to go to 100%, since a 4-fold molar excess of one fragment could not drive the splicing of the second fragment to completion (data not shown). This result suggests that some fraction of each fragment is incapable of splicing.

Five standard protocols were employed that consisted of (i) having one, both or neither fragment in urea prior to the pre-incubation step, (ii) a pre-incubation step at different urea concentrations, and (iii) a splicing step at different urea concentrations (Table I). The initial urea concentration of the separate fragments had no effect on splicing, but the urea concentration in the pre-incubation buffer drastically affected splicing. There was little difference in MP formation when there was 3.0–7.2 M urea in the pre-incubation buffer, but splicing was blocked or inhibited in pre-incubation buffers containing 0–1.8 M urea. No splicing was observed if the pre-incubation mix was diluted immediately into splicing buffer. Splicing efficiency improved with increasing pre-incubation times up to 4 h, after which there was only a small increase in spliced product. Allowing the diluted pre-incubated samples to 'renature' in splicing buffer at 4°C for 0–12 h before shifting to 37°C had no effect on splicing efficiency (data not shown). The presence of ≤ 6.0 M urea in the splicing buffer had little effect on splicing efficiency (Figure 4), but splicing was blocked in 8.0 M urea after 4 h at 37°C (Table I). Use of $I_{C3}P$ or MI_{N3} crude extracts had no significant effect on splicing, indicating that reassembly could occur in the presence of exogenous proteins (data not shown). Varying the pH (5.5 or 7.9) of the pre-incubation and splicing buffers had no effect on splicing of MI_{N3} plus $I_{C3}P$ (data not shown). The effects of several folding aids were tested in the pre-incubation or splicing buffers, or both. Triton X-100 (1%), glycerol (10%), PEG 8000 (0.3%), arginine (0.5 M) and SDS

Fig. 5. Splicing with fragment pairs resulting in gaps or overlaps and with MIP intein deletions. Left panel: splicing of $I_{C1}P$ with MI_{N1} or MI_{N3} using protocol 2 after 0 or 60 min incubations in splicing buffer. Splicing of MI_{N3} plus $I_{C1}P$ (330 amino acid overlap) results in production of MP, I_{N3} and I_{C1} . Middle panel: splicing of MI_{N3} plus $I_{C2}P$ containing an overlap of 190 amino acids. Right panel: fragment pairs resulting in a gap in the intein sequence (MI_{N1} plus either $I_{C3}P$ or $I_{C2}P\Delta$) and a MIP precursor with a deletion of intein residues 109–440 ($MI\Delta 109-440P$) which mimics the gap in the MI_{N1} plus $I_{C3}P$ pair failed to splice or cleave after protocol 2 and incubation in splicing buffer for 120 min. Lane S, molecular weight standards.

Table I. Comparison of pre-incubation and splicing reaction conditions with complementary fragment pairs

Protocol ^a	MI fragment (M urea)	IP fragment (M urea)	Pre-incubation M urea	Pre-incubation time	Splicing reaction M urea	Splicing efficiency ^b
Splicing efficiency versus splicing protocol						
1	MI_{N2} (6.0 M)	$I_{C2}P$ (6.0 M)	6.0 M	4 h	0.6 M	47 ± 3%
2	MI_{N2} (0 M)	$I_{C2}P$ (7.2 M)	3.6 M	4 h	0.36 M	53 ± 9%
2	MI_{N3} (0 M)	$MI_{C3}P$ (0 M)	3.6 M	4 h	0.36 M	54 ± 6%
3	MI_{N2} (6.0 M)	$I_{C2}P$ (6.0 M)	6.0 M	4 h	6.0 M	52 ± 3%
3	MI_{N3} (0 M)	$MI_{C3}P$ (0 M)	8.0 M	4 h	8.0 M	0%
4	MI_{N3} (0 M)	$MI_{C3}P$ (0 M)	0 M	4 h	0 M	0%
Splicing efficiency versus complementary fragment pair						
2	MI_{N1} (0 M)	$I_{C1}P$ (7.2 M)	3.6 M	4 h	0.36 M	0%
2	MI_{N2} (0 M)	$I_{C2}P$ (7.2 M)	3.6 M	4 h	0.36 M	53 ± 9%
2	MI_{N3} (0 M)	$I_{C3}P$ (7.2 M)	3.6 M	4 h	0.36 M	59 ± 6%
2	MI_{N3} (0 M)	$MI_{C3}P$ (0 M)	3.6 M	4 h	0.36 M	54 ± 6%
Splicing efficiency versus time of pre-incubation						
2	MI_{N2} (0 M)	$I_{C2}P$ (7.2 M)	3.6 M	0 h	0.36 M	0%
2	MI_{N2} (0 M)	$I_{C2}P$ (7.2 M)	3.6 M	0.5 h	0.36 M	23 ± 6%
2	MI_{N2} (0 M)	$I_{C2}P$ (7.2 M)	3.6 M	4 h	0.36 M	53 ± 9%
2	MI_{N2} (0 M)	$I_{C2}P$ (7.2 M)	3.6 M	20 h	0.36 M	74 ± 5%
Splicing efficiency versus urea concentration in the pre-incubation reaction						
4	MI_{N3} (0 M)	$MI_{C3}P$ (0 M)	0 M	4 h	0 M	0%
5	MI_{N3} (0 M)	$I_{C3}P$ (7.2 M)	0.9 M	4 h	0.09 M	0%
5	MI_{N3} (0 M)	$I_{C3}P$ (7.2 M)	1.8 M	4 h	0.18 M	9 ± 1%
2	MI_{N3} (0 M)	$I_{C3}P$ (7.2 M)	3.6 M	4 h	0.36 M	59 ± 6%

^aFragments were mixed and pre-incubated at 4°C for the indicated times and urea concentrations. Pre-incubated samples were then diluted into splicing buffer at the indicated final urea concentrations and immediately incubated at 37°C for 2 h.

^bSplicing efficiency = (moles of MP produced/initial moles of limiting substrate) × 100 and was calculated from two or more independent experiments.

(0.01%) had no effect on the rate of splicing or the amount of spliced product observed after a 2 h splicing reaction at 37°C. However, 0.1% SDS in the pre-incubation buffer blocked splicing.

Functional assembly of split inteins reconstitutes PI-PspI endonuclease activity

The MI_{N2} plus IC₂P pair was also tested for endonuclease activity. After a 2 h splicing reaction including pre-treatment in urea (protocol 1), the reassembled and spliced MI_{N2} plus IC₂P sample was added to a standard PI-PspI digestion mixture. The reconstituted intein yielded the same cleavage pattern as the MIP52 control, although in both cases the amount of enzyme added was insufficient to yield a complete digest (Figure 6, lanes 2 and 3). Cleavage was dependent on pre-incubation in urea, since no digestion was observed if the fragment pair was added directly to the DNA digestion mixture (Figure 6, lane 1). These results suggest that the presence of the DNA substrate does not obviate the need for the urea pre-incubation step in fragment assembly. The individual fragments did not have detectable endonuclease activity (Figure 6, lanes 4 and 5) despite the fact that the IC₂P fragment begins 30 amino acids N-terminal to intein block C and contains all of the putative homing endonuclease motifs (intein blocks C, D, E and H) (Mueller *et al.*, 1994; Pietrovski, 1994; Perler *et al.*, 1997a). With the information presently available, it is difficult to correlate the structure of the Sce VMA intein with amino acid sequence although the Sce VMA endonuclease domain is reported to begin 27 amino acids N-terminal to intein block C (Duan *et al.*, 1997). However, based on comparison with the hedgehog processing domain and the analysis of Hall *et al.* (1997), PI-PspI residues 250–538 may not contain the entire endonuclease domain and do not contain the proposed DNA recognition region immediately following block B.

Splicing and cleavage with individual fragments or fragment pairs resulting in gaps

Previous studies indicated that cleavage at either Psp Pol-1 intein splice site does not require the conserved residues at the opposite splice junction (Xu and Perler, 1996). All six individual fragments were therefore tested for the

Fig. 6. Intein reconstitution also re-establishes PI-PspI endonuclease activity. Endonuclease activity of MIP fragments was assayed on pAKR7, which is a 3.7 kb plasmid containing a single PI-TliII (PI-PspI) site. Only *cis*-spliced intein from MIP52 and MI_{N2} plus IC₂P reconstituted using protocol 1 were able to cleave the linearized plasmid into 2.3 and 1.4 kb pieces. Lane 1, MI_{N2} plus IC₂P directly added to the endonuclease reaction without pre-treatment in urea; lane 2, MI_{N2} plus IC₂P after protocol 1 treatment; lane 3, MIP52; lane 4, MI_{N2} after protocol 1 treatment; lane 5, IC₂P after protocol 1 treatment.

ability to induce cleavage at the single splice junction present in that fragment. However, no cleavage was observed with any single fragment under any condition tested (Figure 7 and data not shown).

Several lines of evidence indicate that the splicing domain is limited to the terminal regions of the intein and that an endonuclease domain is inserted between the N- and C-terminal splicing subdomains (Pietrovski, 1994, 1996; Chong and Xu, 1997; Derbyshire *et al.*, 1997; Duan *et al.*, 1997; Hall *et al.*, 1997; Perler *et al.*, 1997a; Telenti *et al.*, 1997). Therefore, fragment pairs resulting in gaps or deletions of intein sequence were tested for the ability to splice. All combinations of fragments that resulted in a gap in the intein (MI_{N1} plus IC₁P, IC₂P or IC₃P and MI_{N2} plus IC₃P) failed to splice or cleave (Figure 5, right panel and data not shown). To eliminate the possibility that the failure to splice was due to a failure to reassociate, MIP deletions were made to mimic these gaps, generating MIΔP precursors missing intein residues 109–440, 150–440, 251–440 or 273–440 and containing a 4–79 amino acid flexible linker at the deletion site. Splicing was performed under standard *cis*-splicing conditions (Xu *et al.*, 1993) or *trans*-splicing conditions. No splicing was observed with any precursor containing a deletion (Figure 5, right panel and data not shown), indicating that the failure of the gapped pairs to splice appears to be unrelated to splicing *in trans*.

Fig. 7. *Trans*-splicing with other exteins. Splicing *in trans* of non-MIP precursors was examined after pre-incubation in 3.6 M urea buffer (protocol 2, +Urea lanes) or without pre-incubation in urea-containing buffers (protocol 4, -Urea lanes) prior to incubation in splicing buffer for 0–60 min at 37°C. Left panel: the C-extein (P) of MI_{C3}P was replaced with the tripeptide, Gly-Leu-Asn (GLN), and MI_{C3}GLN was reacted with MI_{N3}. The single substrate fragments were also incubated by themselves for 60 min (protocol 2). Splicing was assessed by quantifying the appearance of spliced products (MGLN, I_{N3} and MI_{C3}). Note that MI_{C3}GLN and MI_{C3} co-migrate in this gel system and thus MI_{C3}(GLN) represents either or both fragments. Some splicing products (I_{N3} and MGLN) are already observed after overnight pre-incubation in urea (0 min). Right panel: the N-extein (M) of MI_{N3} was replaced with the Lck fragment, L_N. L_NI_{N3} and MI_{C3}GLN were pre-incubated in splicing buffer at 4°C for 30 min without pre-treatment in any urea-containing buffers and then incubated at 37°C for 60 min. The single fragments were also incubated by themselves at 37°C for 60 min. Since the L_NGLN spliced product is too small to be observed in this gel system and since L_N stains poorly with Coomassie blue, intein activity was assessed by the conversion of L_NI_{N3} to I_{N3}, which requires N-terminal cleavage.

Splicing with overlapping fragments or purification tags at the split sites

Previous studies indicated that, in a few cases, split proteins could contain one or more vector-derived amino acids or overlapping redundant sequences at the split site (Matsuyama *et al.*, 1990; Burbaum and Schimmel, 1991). Therefore, all possible pairs resulting in an overlap of intein sequence (MI_{N2} plus $I_{C1}P$, MI_{N3} plus $I_{C2}P$, MI_{N3} plus $I_{C1}P$) were assayed for the ability to splice. All overlapping pairs spliced, even though $I_{C1}P$ failed to splice with its complement, MI_{N1} (data not shown and Figure 5, left and center panels). In the case of MI_{N3} plus $I_{C1}P$, a functional intein was reconstituted despite the presence of 330 amino acids of redundant intein sequence. It is assumed that as the fragment pairs reassociated, redundant sequences were extruded from a location that did not interfere with formation of the splicing active site.

These data suggested that purification tags could be added to split sites. A six residue His tag was added to the split site of MI_{N3} and had no effect on splicing. MBP was fused to the split site of $I_{C3}P$ to yield $MI_{C3}P$ (Figure 2). The presence of an MBP (43 kDa) affinity tag in the I_C fragment, $MI_{C3}P$, had no effect on splicing efficiency (Figure 3 and Table I). However, the addition of MBP to the N-terminus of $I_{C3}P$ converted the insoluble $I_{C3}P$ fragment into a soluble $MI_{C3}P$ fragment. The presence of MBP at the N-terminus of fusion proteins often improves solubility of recombinant proteins in *E.coli*. Despite the fact that both MI_{N3} and $MI_{C3}P$ were synthesized as soluble proteins, no splicing was observed unless the fragments were pre-treated in 3.6 M urea (Figure 3).

Trans-splicing with other exteins

To test the general applicability of the *trans*-splicing system, other exteins were substituted for MBP and paramyosin in the MI_{N3} plus $MI_{C3}P$ system. In L_NI_{N3} , the N-extein (M) of MI_{N3} was replaced by a 9 kDa fragment of Lck tyrosine kinase encoding residues 52–121 (Perlmutter *et al.*, 1988). L_NI_{N3} plus either $MI_{C3}P$ or $I_{C3}P$ spliced after pre-incubation in 3.6 M urea (data not shown). In $I_{C3}L_C$, the paramyosin domain of $I_{C3}P$ was replaced by a 10 kDa fragment of Lck tyrosine kinase encoding residues 122–226 (Perlmutter *et al.*, 1988). Both MI_{N3} plus $I_{C3}L_C$ or L_NI_{N3} plus $I_{C3}L_C$ were able to splice if pre-treated in 3.6 M urea (data not shown). The paramyosin domain of $MI_{C3}P$ was then replaced by three amino acids (Gly–Leu–Asn or GLN), yielding $MI_{C3}GLN$. When $MI_{C3}GLN$ was mixed with a complementary fragment, MI_{N3} , splicing products $MGLN$, I_{N3} and MI_{C3} were observed only after pre-treatment in 3.6 M urea (Figure 7, left panel). Note that MI_{C3} and $MI_{C3}GLN$ are indistinguishable on these gels.

Two complementary fragments containing the smallest exteins, L_NI_{N3} and $MI_{C3}GLN$, were mixed directly in splicing buffer (Figure 7, right panel). Since the spliced exteins (LGLN) are too small to be clearly observed and since the Lck fragment stains poorly with Coomassie blue, reactions were scored positive for intein activity (splicing or cleavage) if L_NI_{N3} was converted to I_{N3} with time. As a control, L_NI_{N3} and $MI_{C3}GLN$ were also incubated as above in the absence of the complementary fragment. I_{N3} only accumulated when L_NI_{N3} was mixed with $MI_{C3}GLN$, indicating that the intein was active without pre-treatment

in urea. However, the reaction under native conditions was not very efficient.

Discussion

Precursor fragments split within the Psp Pol-1 intein can be synthesized in separate hosts and then assembled *in vitro* to generate a fully active intein with both protein splicing and endonuclease activities. Remarkably, the intein fragments were able to assemble at temperatures of up to 100°C below their normal synthesis temperature. Once reconstituted, the active intein directed ligation of several test exteins including the chimera MP and L_N plus L_C , a Lck tyrosine kinase fragment spanning amino acids 52–226. *Trans*-splicing time courses confirmed the protein splicing pathway (Xu and Perler, 1996) since N-terminal and C-terminal cleavage could be monitored by the release of distinguishable intein fragments. Two of the three intein fragmentation sites yielded complementary precursor fragments that were capable of splicing, which is similar to previous data with other split proteins where only a fraction of the complementary pairs are able to reassemble (Matsuyama *et al.*, 1990; Burbaum and Schimmel, 1991). All combinations of overlapping fragments were also able to assemble into an active intein, including pairs with 330 amino acids of redundant intein sequence, requiring sufficient flexibility to displace these extra residues while forming the functional intein. The ability of these fragments to accommodate overlapping sequences suggested that they might be able to accommodate affinity tags at the split sites. The addition of a His tag to the C-terminus of I_N fragments had no effect on splicing, nor did the presence of a 43 kDa MBP affinity tag at the split site of $I_{C3}P$ have any effect on splicing, although it converted a previously insoluble fragment into a soluble fragment. No fragment pair resulting in a gap in the intein sequence and no MIP precursor containing a deletion mimicking these gaps was able to splice, suggesting that the failure of the gap pairs to splice is unrelated to the fragment assembly process.

Expression of MIP precursors is often accompanied by some degree of N-terminal or C-terminal cleavage *in vivo* (Xu *et al.*, 1993; Xu and Perler, 1996), but no *in vivo* or *in vitro* cleavage was observed with any individual fragment. This was unexpected since previous studies (Xu and Perler, 1996) had indicated that conserved Psp Pol-1 intein splice junction residues at one splice site are not required for cleavage at the opposite site and since even the smallest fragments contain all of the putative N-terminal (blocks A and B) or C-terminal (blocks F and G) splicing motifs (Petrokovski, 1996; Duan *et al.*, 1997; Hall *et al.*, 1997; Perler *et al.*, 1997a; Telenti *et al.*, 1997). These results suggest that residues from both the N- and C-terminal subdomains of the Psp Pol-1 intein are required for cleavage at both splice junctions. This hypothesis is supported by the structures of the Sce VMA intein and hedgehog protein autoprocessing domain which indicate that the N- and C-terminal intein subdomains fold together to form one intermingled splicing domain.

This study also examined the effect of urea on the splicing reaction. Urea was required to solubilize the insoluble I_CP fragments. With all but the smallest exteins, pre-incubation in urea buffers was required for reconstitu-

tion of intein activity. Up to 6.0 M urea in the splicing buffer had no effect on splicing efficiency, but splicing was inhibited in 8.0 M urea. Splicing requires bringing at least three separate intein regions together to form the splicing active site (Ser1, His96, Asn537 and Ser538 in Psp Pol-1 intein) and splicing *in vitro* with MIP is very slow (half-time of 20–30 min, Xu and Perler, 1996). These results therefore suggest that urea concentrations of 6.0 M or less are not fully denaturing the intein since it is unlikely that a fully denatured intein could bring the above residues into proximity long enough for the slow *in vitro* splicing reaction to occur. These results suggest that the mechanisms resulting in thermostability of this protein from an extreme thermophile may also make it resistant to other forms of denaturation such as urea. In the future, it should be possible to maximize splicing of a split intein precursor containing any extein by determining optimum conditions for fragment solubility and precursor assembly. However, since the presence of the intein may prevent proper folding of the foreign extein protein, it may have to be denatured and renatured after splicing.

The cleavage and reassembly of protein splicing precursors opens up new avenues of protein analysis. Although the reconstituted precursor contains a break in the peptide backbone in the intein domain, after splicing the exteins are covalently linked with a native peptide bond (Cooper *et al.*, 1993). *Trans*-splicing of split precursors can thus be used to label or modify only a portion of the intact extein product. For example, an N-terminal fragment can be isolated from *E. coli* grown in media enriched with ^{13}C or ^{15}N . After splicing, the intact protein would only be labeled in the region of the N-extein. Such a partially labeled protein could be used to simplify structural determination by NMR analysis or possibly to allow the determination of larger protein structures. Another potential use of splicing with split inteins would be the modification by glycosylation, phosphorylation, dephosphorylation, etc. of only a subset of sites in a protein to determine which post-translationally modified site is important for enzyme activity. Finally, *trans*-splicing allows overexpression of highly toxic proteins, since no single cell contains the entire toxic protein.

Materials and methods

General procedures

Protein samples were electrophoresed on either 4–20% SDS–PAGE gels (Daiichi) or 12% SDS–PAGE gels (Novex). Protein concentrations were determined by the Bradford assay (Bio-Rad). Western blots were probed with anti-MBP, anti-*D. immitis* paramyosin or anti-Psp Pol-1 intein sera as previously described (Xu *et al.*, 1993). Western blot data are not shown, but were used to confirm the composition of all splicing and cleavage products. Protein Marker, Broad Range (New England Biolabs) standards were used. All cloning enzymes and oligonucleotides were from New England Biolabs and were used according to the manufacturer's instructions. All PCR fragments were sequenced in both directions by the New England Biolabs DNA sequencing core facility.

Construction of pMI_{N1}, pMI_{N2} and pMI_{N3}

DNA fragments encoding all or part of MBP and the indicated N-terminal fragments of the Psp Pol-1 intein were synthesized by PCR from pMIP21 (Figure 2) (Xu *et al.*, 1993). In each case, a stop codon and restriction enzyme site were introduced after the last intein codon. However, no extra residues were present at the split site, unless added later as an affinity tag. The fragment containing MBP and the first 249 codons of the intein was ligated into pAII17 (Hodges *et al.*, 1992)

yielding pMI_{N2}, and fragments containing the first 108 or 440 intein codons were ligated into *Xho*I–*Bam*HI-digested pMI_{N2} yielding pMI_{N1} and pMI_{N3}, respectively. The primers were: 5'-GGAATTCATATGAA-AATCGAAGAAGGT-3' (pMI_{N2}); NEB 1237 (pMI_{N1} and pMI_{N3}), 5'-GGTCGTGACACTGTCGATGAAGCC-3', (pMI_{N1}), 5'-GGGGATCCTTACTCAACGAGATCCCCGTCTCTAT-3'; (pMI_{N2}), 5'-CGGGATCCCCGTATAGTGAGATAACGTCCCG-3'; and (pMI_{N3}), 5'-ATTGGATCCTTATCTGTATTCCGTAACCTTA-3'. PCR mixtures contained Vent DNA polymerase buffer (New England Biolabs), 0.2 mM each dNTP, 0.4 μM primers, 100 ng of plasmid DNA and 1 U of Vent DNA polymerase (New England Biolabs) in a 100 μl reaction. Amplification was carried out using a Perkin-Elmer Cetus 480 thermal cycler at 94°C for 30 s, 52°C for 30 s and 72°C for 135 s for 15–17 cycles. A C-terminal six amino acid His tag was added to the split site of MI_{N3} and MI_{N2} by standard procedures.

Construction of pI_{C1}P, pI_{C2}P, pI_{C2}Δ, pI_{C3}P and pMI_{C3}P

Clones containing C-terminal fragments beginning at intein amino acids Thr110 (pI_{C1}P), Met250 (pI_{C2}P) or Lys441 (pI_{C3}P) were constructed by PCR from pMIP21 as described above (Figure 2). The I_{C2}P fragment was subcloned into a pAII17 derivative, yielding a six residue His tag at the end of the paramyosin gene. The remaining two clones were generated by replacing the I_{C2} sequence in *Nde*I–*Bam*HI-digested pI_{C2}P with PCR products. The primers used were 5'-GGGCATATGACTGGG-GAGGATGTCAAAATT-3' (pI_{C1}P); 5'-GGAATTCATATGCCAGA-GGAAGAACTG-3' (pI_{C2}P); 5'-GAACATATGAAGAAAAAGAA-TGTATATCACTCTC-3' (pI_{C3}P); 5'-ATAGTTTAGCGGCCGCTCAC-GACGTTGTAAAACG-3' (pI_{C2}P); and 5'-GGGGATCCAAAGCCAGCAAGGAAATTCTC-3' (pI_{C1}P and pI_{C3}P). An 11 kDa (118 codon) C-terminal deletion was generated in I_{C2}Δ to ease analysis since MI_{N2} and MP co-migrate on SDS–PAGE; this deletion had no effect on splicing. pI_{C2}P was digested with *Nsi*I and *Sal*I, and blunted with T4 DNA polymerase prior to ligation. An *Nde*I–*Pst*I fragment from pI_{C3}P (encoding I_{C3}P) was cloned into the *Eco*RI–*Pst*I site of pMAL-c2 (New England Biolabs) to form an in-frame fusion of MBP with I_{C3}P, yielding pMI_{C3}P.

Construction of L_{N1N3}, MI_{C3}L_C and MI_{C3}GLN

The L_{N1N3} Lck tyrosine kinase fusion, encoding Lck residues 52–121, was produced by subcloning the PCR product amplified from the Lck cDNA clone, pCDNA1 (Perlmutter *et al.*, 1988), in place of MBP in pMI_{N3}His. PCR primers were 5'-GCTTACGCATATGGGCTCCAATC-CGCCGGCT-3' and 5'-AGTGGTACCCATTCTTCCGGTAAAAATGCTGTTTCGTTTGGCCACAAA-3'. The *Nde*I–*Kpn*I-digested PCR products were gel purified and ligated with pMI_{N3} digested with the same enzymes.

The L_C fragment encoding Lck tyrosine kinase residues 122–226 was generated as above by PCR using primers 5'-GCGGATCCCTCTATGCACATAATAGCCTGGAGCCCGAACCC-3' and 5'-GGGCGAAGCTTACTGGCAGGGGCGGC-3'. To replace paramyosin with L_C, the PCR product and pMI_{C3}P were digested with *Bam*HI–*Hind*III and ligated as above. To replace paramyosin with the tripeptide Gly–Leu–Asn, pMI_{C3}P was digested with *Bam*HI–*Hind*III, and a double-stranded oligonucleotide cassette (5'-GATCCCTCTATGCACATAATTAGGCCTCAATTAA-3' and 5'-AGCTTTAATTGAGGCCTGAATTATGTGCATAGAGG-3') was ligated to the cut plasmid as above.

Expression and purification

pMI_{N1}, pMI_{N2}, pMI_{N3}, pMI_{C3}P, pMI_{C3}L_C and pMI_{C3}GLN were transformed into either ER2520 (*E. coli* B F-IDE3 (=1 *sBam*HI Δ EcoRI-B *int::lacI::PlacUV5::T7 gene1 imm21 Δ nin5 Δ (*mcrC-mrr*)102::Tn10 gal ompT [lon, dcm] [pLysS: CmR orip15A T7 gene 2]) or ER2538 (*E. coli* B; F-IDE3(=1 *sBam*HI Δ EcoRI-B *int::lacI::PlacUV5::T7 gene1 imm21 Δ nin5 Δ fluA2 [lon] ompT gal sulA11 Δ (*mcrC-mrr*)114::IS10 R(*mcr-73::miniTn10-TetS*)2 *endA1 R(zgb210::Tn10-TetS)* [pLysS: CmR orip15A T7 gene 2]) (Elisabeth Raleigh, New England Biolabs) and grown at 30°C in LB medium plus 100 $\mu\text{g}/\text{ml}$ of ampicillin to an OD₆₀₀ of ~0.5. The culture was induced with 0.4 mM isopropyl- β -D-thiogalactopyranoside (IPTG) and incubated overnight. The cells were sonicated in 50 ml of amylose column buffer (20 mM NaPO₄, pH 8.0, 0.5 M NaCl, 1.0 mM Na₂EDTA) and purified over an amylose column as described by the manufacturer (New England Biolabs).**

pI_{C2}P, pI_{C2}Δ, pI_{C1}P, pI_{C3}P and pL_{N1N3} were transformed into ER2520 and induced as above, except pI_{C1}P was induced with 0.04 mM IPTG for 2 h (I_{C1}P). Since a His tag was present after the paramyosin domain or the C-terminus of L_{N1N3}, frozen cells were lysed by sonication in Ni²⁺ binding buffer (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 16 mM

imidazole) and centrifuged at 10 000 g for 30 min. Insoluble I_CP proteins were solubilized in Ni²⁺ binding buffer containing 6.0 M urea and purified over a Ni²⁺ charged column in the presence of 6.0 M urea (Novagen, 5 ml of resin) as described by the manufacturer. Alternatively, insoluble proteins were purified by merely washing the pellet several times with Ni²⁺ binding buffer containing 1.0 M urea prior to resuspension in 6.0 M urea.

Fragment assembly (trans-splicing) protocols and quantitation

Coomassie blue-stained gels were digitized with a Microtek Scanmaker III and analyzed with NIH Image 1.51 software for quantitation. Samples were not normalized for protein loss during the time course since there appeared to be unequal loss of each substrate and product. However, it should be noted that there was up to 35% loss of total protein during most splicing reactions of 1.5 h or more, indicating precipitation during the reaction, especially of I_C-containing fragments which are known to precipitate with time in the absence of urea. There may also be differences in staining of fragments with Coomassie blue since splicing of MIP *in cis* yields only 71% as many moles of MP product as I product, with no compensating increase in M or P cleavage products. Therefore, splicing efficiency was calculated to reflect actual yields of spliced product (the product of interest) by dividing the moles of spliced product at the end of the reaction by the initial moles of limiting substrate (usually MI_N) and multiplying by 100.

Purified protein fragments were stored in either amylose elution buffer, Ni²⁺ elution buffer (± 6.0 M urea) or exchanged into buffer E [50 mM Tris-HCl pH 7.5, 5% acetate 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 140 mM β-mercaptoethanol and 7.2 M urea, equilibrated to pH 7.5]. The proteins were then combined to a final concentration of 0.5–2.5 mg/ml. Several pre-incubation protocols were examined involving treatment alone or after mixing in various concentrations of urea (0–8.0 M) and different buffers for differing times at 4 or 37°C prior to dilution (1:10 to 1:50) into splicing buffer (20 mM NaPO₄, pH 6, 0.5 M NaCl, 1 mM EDTA). However, most experiments presented in this study were performed with one of the following standard protocols (Table I). In protocol 1, both fragments were in buffer E (7.2 M urea) or Ni²⁺ elution buffer (6.0 M urea) prior to mixing. In protocol 2, MI_N in amylose elution buffer was combined with an equal volume of C-terminal fragment in buffer E or Ni²⁺ elution buffer, resulting in a 3.6 or 3.0 M urea pre-incubation buffer, respectively. In protocol 3, fragments were mixed in various concentrations of urea (0–8.0 M) and then diluted into splicing buffer containing the same concentration of urea. In protocol 4, the fragments were mixed in amylose elution buffer without urea and diluted into splicing buffer without urea. In protocol 5, MI_N in amylose elution buffer was combined with I_CP in buffer E (7.2 M urea) and varying amounts of amylose elution buffer to yield pre-incubation buffers containing 0–3.6 M urea and then diluted 10-fold into splicing buffer without urea. All fragment mixtures except those using protocol 4 were pre-incubated in urea-containing buffers for 0 h to overnight at 4°C. Following this pre-incubation step, the mixtures were diluted 10- to 50-fold in splicing buffer and immediately incubated at 37°C for 0–24 h to stimulate the splicing reaction. Most experiments presented were diluted 10-fold since no significant difference was observed with higher dilutions. Purified I_C1P, I_C2P, I_C3P and I_C3LC precipitated within 2–4 h after rapid dilution out of urea-containing buffers. Therefore, in most cases, splicing reactions were limited to 2 h to avoid loss of these substrates due to precipitation rather than splicing.

In some experiments, the following folding aids were added to either or both the pre-incubation or splicing buffers (protocol 2): Triton X-100 (1%), glycerol (10%), PEG 8000 (3 mg/ml), arginine (0.5 M) and SDS (1 mg/ml and 0.1 mg/ml).

Endonuclease assays

pAKR7 contains the 714 bp *Eco*RI fragment from pAKK4 (Hodges *et al.*, 1992) in Bluescript SK– and encodes a single PI-TliII recognition site created by the deletion of the *Tli pol-I* intein from the *Thermococcus litoralis* DNA polymerase gene. This site is similar, but not identical, to the PI-PspI recognition site predicted by deletion of the *Psp pol-I* intein. Although PI-TliII and PI-PspI are isoschizomers (F.Perler, unpublished data), it is not known whether they cut each other's recognition site with the same efficiency. Digestions were performed for 1 h at 50°C in 0.1 M NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 8.6 (at 25°C). Similar amounts of PI-PspI were present in each reaction in the form of (i) MIP52 (>80% present as the MIP precursor at the beginning of the reaction) which contains an insert of MILVA prior to Ser1 of MIP, (ii) single fragments MI_N2 or I_C2P, (iii) MI_N2 and I_C2P fragment

pairs that had been pre-assembled using protocol 1 with a 2 h splicing reaction or (iv) MI_N2 and I_C2P directly added to the digestion mixture. The 3.7 kb pAKR7 DNA was linearized with *Xmn*I prior to digestion with PI-PspI so that digestion with PI-PspI would yield fragments of 1.4 and 2.3 kb.

Construction of intein deletions

The insert in pMI_N1 was amplified by PCR from the *Nde*I site at the beginning of the *malE* gene to the end of I_N1 with the addition of an *Spe*I and *Nde*I site at the 3' end. This fragment was digested with *Nde*I and subcloned into the *Nde*I site of pI_C3P to create pMIA109–440P which contains a *Spe*I and *Nde*I site between the intein sequences resulting in the insertion of four amino acids (Thr–Ser–His–Met). Two complementary primers (5'-CT AGG GGC TAT GAC CTG CCC ATG GTT GAG GAA GGA GAG CCT GAC C-3' and 3'-C CCG ATA CTG GAC GGG TAC CAA CTC CTT CCT CTC GGA CTG GGA TC-5') were ligated into *Spe*I-digested pMIA109–440P, resulting in the insertion of 15 amino acids at the deletion site per copy of double-stranded linker. The linker can be inserted in either direction, resulting in different sequence combinations. All subsequent MIP deletion clones were made by digesting pMIA109–440P with *Eco*RI–*Spe*I and substituting similarly digested PCR products. MIP deletion precursors were purified as described above for MI_N proteins. MIA109–440P precursors containing 1–5 copies of the linker, MIA150–440P and MIA273–440P precursors containing 1–3 copies of the linker, and MIA251–440P precursors containing one or two linkers were all tested for the ability to splice under standard *cis*- (Xu and Perler, 1996) and *trans*-splicing conditions.

Acknowledgements

We gratefully acknowledge Jonathan Lee of Boston University for support of R.A.K. and for helpful discussions. We thank Bill Jack for the gift of pAKR7 DNA and Jack Benner and Ted Davis for sharing the results of their Psp Pol-1 intein protease digestion studies. We thank Chris Noren for helpful discussions and Don Comb for support and encouragement. Sections of this study were performed at New England Biolabs as part of the thesis requirements of E.A. and D.P. at the Ecole Supérieure de Biotechnologie de Strasbourg, France (1994–1995).

References

- Anfinsen, C.B. and Scheraga, H.A. (1975) Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.*, **29**, 205–300.
- Bremer, M., Gimble, F.S., Thorner, J. and Smith, C. (1992) VDE endonuclease cleaves *Saccharomyces cerevisiae* genomic DNA at a single site: physical mapping of the VMA1 gene. *Nucleic Acids Res.*, **20**, 5484.
- Burbaum, J.J. and Schimmel, P. (1991) Assembly of a class I tRNA synthetase from products of an artificially split gene. *Biochemistry*, **30**, 319–324.
- Chong, S. and Xu, M.Q. (1997) Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs. *J. Biol. Chem.*, **272**, 15587–15590.
- Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F.B. and Xu, M.Q. (1996) 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**, 22159–22168.
- Cooper, A.A., Chen, Y., Lindorfer, M.A. and Stevens, T.H. (1993) Protein splicing of the yeast *TFPI* intervening protein sequence: a model for self-excision. *EMBO J.*, **12**, 2575–2583.
- Derbyshire, V., Wood, D.W., Wu, W., Dansereau, J.T., Dalgaard, J.Z. and Belfort, M. (1997) 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**, 11466–11471.
- Duan, X., Gimble, F.S. and Quijoch, F.A. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*, **89**, 555–564.
- Gross, M., Wyss, M., Furter-Graves, E.M., Wallimann, T. and Furter, R. (1996) Reconstitution of active octameric mitochondrial creatine kinase from two genetically engineered fragments. *Protein Sci.*, **5**, 320–330.
- Hall, T.M.T., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A. and Leahy, D.J. (1997) Crystal structure of a hedgehog autoprocessing domain: conservation of structure, sequence and cleavage mechanism between hedgehog and self-splicing proteins. *Cell*, **91**, 85–97.

- Hodges,R.A., Perler,F.B., Noren,C.J. and Jack,W.E. (1992) Protein splicing removes intervening sequences in an archaea DNA polymerase. *Nucleic Acids Res.*, **20**, 6153–6157.
- Kanaya,E. and Kanaya,S. (1995) Reconstitution of *Escherichia coli* RNase HI from the N-fragment with high helicity and the C-fragment with a disordered structure. *J. Biol. Chem.*, **270**, 19853–19860.
- Kato,I. and Anfinsen,C.B. (1969) On the stabilization of ribonuclease S-protein by ribonuclease S- peptide. *J. Biol. Chem.*, **244**, 1004–1007.
- Kawasaki,M., Nogami,S., Satow,Y., Ohya,Y. and Anraku,Y. (1997) Identification of three core regions essential for protein splicing of the yeast Vma1 protozyme. *J. Biol. Chem.*, **272**, 15668–15674.
- Matsuyama,S., Kimura,E. and Mizushima,S. (1990) Complementation of two overlapping fragments of SecA, a protein translocation ATPase of *Escherichia coli*, allows ATP binding to its amino-terminal region. *J. Biol. Chem.*, **265**, 8760–8765.
- Mueller,J.E., Bryk,M., Loizos,N. and Belfort,M. (1994) Homing endonucleases. In Linn,S.M. Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 111–143.
- Perler,F.B. *et al.* (1992) Intervening sequences in an Archaea DNA polymerase gene. *Proc. Natl Acad. Sci. USA*, **89**, 5577–5581.
- Perler,F.B., Davis,E.O., Dean,G.E., Gimble,F.S., Jack,W.E., Neff,N., Noren,C.J., Thorner,J. and Belfort,M. (1994) Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucleic Acids Res.*, **22**, 1125–1127.
- Perler,F.B., Olsen,G.J. and Adam,E. (1997a) Compilation and analysis of intein sequences. *Nucleic Acids Res.*, **25**, 1087–1093.
- Perler,F.B., Xu,M.-Q. and Paulus,H. (1997b) Protein splicing and autoprolysis mechanisms. *Curr. Opin. Chem. Biol.*, **1**, 292–299.
- Perlmutter,R.M., Marth,J.D., Lewis,D.B., Peet,R., Ziegler,S.F. and Wilson,C.B. (1988) Structure and expression of lck transcripts in human lymphoid cells. *J. Cell. Biochem.*, **38**, 117–126.
- Petrokovski,S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci.*, **3**, 2340–2350.
- Petrokovski,S. (1996) A new intein in Cyanobacteria and its significance for the spread of inteins. *Trends Genet.*, **12**, 287–288.
- Rost,B., Sander,C. and Schneider,R. (1994) PHD—an automatic mail server for protein secondary structure prediction. *CABIOS*, **10**, 53–60.
- Sancho,J. and Fersht,A.R. (1992) Dissection of an enzyme by protein engineering. The N and C-terminal fragments of barnase form a native-like complex with restored enzymatic activity. *J. Mol. Biol.*, **224**, 741–747.
- Shao,Y., Xu,M.Q. and Paulus,H. (1995) Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence. *Biochemistry*, **34**, 10844–10850.
- Shao,Y., Xu,M.-Q. and Paulus,H. (1996) Protein splicing: evidence for an N–O acyl rearrangement as the initial step in the splicing process. *Biochemistry*, **35**, 3810–3815.
- Tasayco,M.L. and Chao,K. (1995) NMR study of the reconstitution of the beta-sheet of thioredoxin by fragment complementation. *Proteins*, **22**, 41–44.
- Telenti,A., Southworth,M., Alcaide,F., Daugelat,S., Jacobs,W.R. and Perler,F.B. (1997) The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. *J. Bacteriol.*, **179**, 6378–6382.
- Xu,M. and Perler,F.B. (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J.*, **15**, 5146–5153.
- Xu,M., Southworth,M.W., Mersha,F.B., Hornstra,L.J. and Perler,F.B. (1993) *In vitro* protein splicing of purified precursor and the identification of a branched intermediate. *Cell*, **75**, 1371–1377.
- Xu,M., Comb,D.G., Paulus,H., Noren,C.J., Shao,Y. and Perler,F.B. (1994) Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *EMBO J.*, **13**, 5517–5522.

Received August 18, 1997; revised November 27, 1997;
accepted December 17, 1997