The EMBO Journal vol.15 no.19 pp.5146–5153, 1996

The mechanism of protein splicing and its modulation by mutation

Ming-Qun Xu and Francine B.Perler

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

Protein splicing results in the expression of two mature proteins from a single gene. After synthesis of a precursor protein, an internal segment (the intein) is excised and the external domains are joined together. A self-catalyzed mechanism for this cleavage–ligation reaction is presented, based on mutagenesis data and analysis of splicing intermediates. Mutations were used to block various steps in the protein splicing pathway, allowing each isolated step to be studied independently. A linear ester intermediate was identified and functional roles for the four conserved splice junction residues were determined. Understanding the mechanism of protein splicing provides a basis for protein engineering studies. For example, inteins can be constructed which fail to splice, but instead cleave the peptide bond at a chosen splice junction.

Keywords: hydroxylamine/intein/protein engineering/protein splicing/thermophile

Introduction

The discovery of protein splicing fundamentally changed our view of the flow of genetic information from DNA to protein (Hirata *et al.*, 1990; Kane *et al.*, 1990; Davis *et al.*, 1991, 1994; Perler *et al.*, 1992; Gu *et al.*, 1993). After synthesis of a precursor protein, an internal protein domain (termed the intein) is excised precisely from the precursor and the two external domains (termed exteins) are ligated together forming a native peptide bond between the exteins (for nomenclature, see Perler *et al.*, 1994). Thus, two proteins are derived from a single gene, one of which is not collinear with the original open reading frame. Extein ligation differentiates protein splicing from other forms of gene expression that result in multiple proteins from a single gene, such as autocleavage or proteolytic processing. Inteins have been found in DNA polymerases, ATPases and RecA proteins. How can such diverse proteins act as protease–ligases? The answer lies in the observation that the intein plus the first C-extein residue contain sufficient information to direct protein splicing, even when the intein is present in a foreign protein context (Davis *et al.*, 1992, 1994; Cooper *et al.*, 1993; Xu *et al.*, 1993).

Not only do inteins mediate the rearrangement of genetic information at the protein level, but they also mediate intein gene mobility (Gimble and Thorner, 1992). Inteins are similar in sequence and function to homing endonucleases present in mobile introns (Lambowitz and Belfort, 1993). Many inteins have been shown to make double strand breaks at the intein insertion site in alleles lacking inteins (Gimble and Thorner, 1992; Perler *et al.*, 1992). Although endonuclease activity is required for intein gene mobility, it is not required for protein splicing (Hodges *et al.*, 1992).

Intein sequences lack similarity except for the two characteristic homing endonuclease dodecapeptide motifs, a small block (block B) containing a conserved His, and four highly conserved splice junction residues (Pietrokovski, 1994). These conserved residues are a Cys or Ser at the beginning of the intein, His-Asn at the end of the intein, and Ser, Thr or Cys at the beginning of the C-extein. Ser, Thr and Cys are functionally similar in that they have side chain hydroxyl or thiol groups which can act as nucleophiles. All protein splicing mechanisms suggest that the catalytic core is formed by bringing the conserved amino acids at the two splice junctions into proximity by proper folding of the intein (Perler *et al.*, 1992, 1994; Cooper and Stevens, 1993; Wallace 1993; Xu *et al.*, 1993, 1994; Clarke, 1994). To understand the basis of protein splicing, we need to unravel the chemical reactions involved and the sequence of these reactions. It is crucial to explain why certain amino acids are highly conserved and to elucidate their chemical roles.

Previous studies with the MIP *in vitro* splicing system have revealed multiple steps in the protein splicing pathway, including a branched protein intermediate (Xu *et al.*, 1993, 1994; Shao *et al.*, 1995). MIP contains the intein from the DNA polymerase of an extremely thermophilic archaeon, *Pyrococcus species* GB-D (I or Psp pol intein-1) inserted between the *Escherichia coli* maltose binding protein (MBP or M, the N-extein) and a fragment of *Dirofilaria immitis* Paramyosin (P, the C-extein). The branched intermediate consists of the N-extein (M) linked at its C-terminus by an alkali-labile bond to the side chain of an amino acid in IP. The branch is resolved by cyclization of the conserved intein C-terminal Asn.

This paper describes mutagenesis and chemical analyses that define the requirements of each step in the splicing pathway and differentiate between previously proposed models. Mutations were used to block different steps in the pathway, defining the conserved residues required for each step and facilitating the study of the mechanism of these isolated steps.

Results

Mutation of the four conserved splice junction residues

pMIP-21, encoding wild-type MIP, allows simple splice junction mutagenesis by cassette replacement. Mutant MIP clones were expressed under conditions which inhibit splicing *in vivo* and during purification. The purified MIP samples were then incubated under conditions which

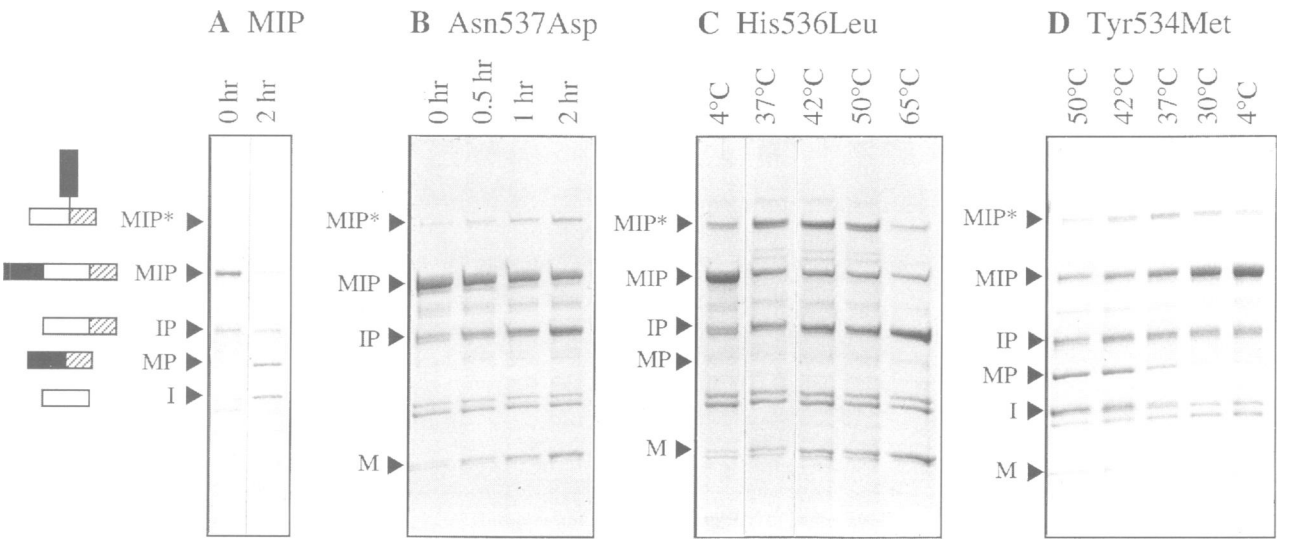

Fig. 1. The effect of mutation on the protein splicing pathway. Wild-type MIP or mutant MIP proteins were purified by affinity chromatography on amylose resin. Depending on the mutation, one of several MIP products co-purified or was formed after incubation in splicing buffer. These products included the branched intermediate (MIP*), the linear precursor (MIP), N-terminal splice junction cleavage products (M + IP), C-terminal splice junction cleavage products (MI + P) or spliced products (MP + I). Samples incubated at 4°C were identical to the initial products isolated. (**A**) Wild-type MIP was incubated at 37°C for 0 or 2 h. (**B**) The Asn537Asp mutant was incubated at 37°C for 0, 0.5, 1 or 2 h resulting in accumulation of the branched intermediate and N-terminal cleavage products with time. (**C**) His536Leu mutant was incubated for 1 h at 4, 37, 42, 50 and 65°C. The branched intermediate accumulated *in vivo* and accumulated further after incubation *in vitro*. (**D**) Tyr534Met mutant was incubated for 2 h at 50, 42, 37, 30 or 4°C, demonstrating reduced splicing at lower temperatures compared with wild-type MIP (A). The black box represents the N-extein (M), the white box represents the intein and the striped box represents the C-extein (P).

Table I. Splicing and cleavage patterns after mutagenesis

N-extein/**Ser1**Ile2.....Ala535**His536Asn537/Ser538**.....(C-extein)

Residue	Mutation	Splices	Branch	N-cleavage	C-cleavage
MIP	none	+	+	+[a]	+[a]
Ser1	Ala	–	–	–	+
	Cys.Thr	±	±	+	+
Ile2	Lys	–	–	–	+
Leu533	Met	+	+	+[a]	+[a]
Tyr534	Met	+[b]	+	+[a]	+[a]
Ala535	Lys	–	–	–	+
Ala535Ser538	LysAla	–	–	–	+
His536	Leu.Asn.Gln	–	+[c]	+	–
Asn537	Asp.Gln	–	+	+	–
Ser538	Ala.Gly	–	–	+	+
	Cys.Thr	±	±	+	+
His536Asn537	LeuAla	–	–	+	–
His536Asn537Ser538	LeuAla.Ala	–	–	+	–
Ser1Asn537	AlaAla	–	–	–	–

The structure of MIP is depicted above the table with the four conserved junction residues in bold and splice junctions indicated by a solidus (/). Psp pol intein-1 is 537 amino acids. The pattern of products observed for each mutation after splicing at 37°C is listed. N-cleavage and C-cleavage indicate products resulting from cleavage at the N- or C-terminal splice junction, respectively; splice, indicates that splicing occurs; branch, indicates the presence of the branched intermediate.

[a]Minor product of *in vivo* cleavage, little increase *in vitro*.

[b]Cold-sensitive, does not splice below 30°C.

[c]Branched intermediate accumulates.

stimulate splicing *in vitro* to study the splicing and/or cleavage reactions (Figure 1A and Xu *et al.*, 1993). The products observed in these MIP mutants included spliced MP + I, the branched intermediate, N-terminal cleavage products (M + IP) and C-terminal cleavage products (MI + P). In Table I, mutations were scored for their effects on splicing, branch formation and cleavage at each splice junction to determine the residues required for each of these steps in the protein splicing pathway. The rates of formation of the major products are listed in Table II.

Replacement of Ser1 in MIP by Cys, Thr or Ala inhibited splicing and N-terminal cleavage despite the conservative nature of some of these substitutions (Tables I and II). The Ser1Cys mutant yielded predominantly C-terminal cleavage products (MI + P), as well as a small amount of spliced product (<5%), branched intermediate and N-terminal cleavage products (M + IP). However, splicing of the Ser1Cys precursor could be stimulated by incubation at higher temperatures (50–65°C). The conservative substitution of Ser1 with Thr completely

Table II. Reaction rates for MIP mutants

Mutation	Major product	Temperature (°C)	Half-time (h)
MIP	MP + I	37	0.5
Ser1Cys	MI + P	37	4
Ser1Ala	MI + P	37	16[a]
His536Leu	MIP*	50	0.5
His536Leu[b]	M + IP	50	3
Ser538Cys	M + I + P	50	2
Ser538Thr	M + I + P	50	2
Ser538Ala	M + I + P	37	4
His536Leu/Ser538Ala	M + IP	37	3.1
Ser1Cys/His536Leu/Ser538Ala	M + IP	37	3.1
His536Leu/Asn537Ala/Ser538Ala	M + IP	37	16
Ala535Lys	MI + P	37	4
Tyr534Met	MP + I	30	–[c]
Tyr534Met	MP + I	42	2
Tyr534Met	MP + I	50	<2[d]

The major products produced after incubation of each mutant MIP precursor are indicated followed by the temperature of the reaction and the number of hours required for 50% of the substrate to be converted to product ($t_{1/2}$). MP + I indicates splicing; MIP* indicates accumulation of branched intermediate; MI + P, indicates C-terminal cleavage; M + IP, indicates N-terminal cleavage; M + I + P, indicates cleavage at both splice junctions.

[a]<10% cleaved by 16 h.

[b]N-terminal cleavage products accumulate after the initial accumulation of the branched intermediate.

[c]<10% spliced by 2 h.

[d]>75% spliced by 2 h.

blocked splicing and inhibited formation of the branched intermediate, resulting in cleavage at both splice junctions. Substitution of Ser1 by Ala completely blocked N-terminal cleavage, branch intermediate formation and splicing, resulting in cleavage at the intein C-terminus (Table I).

His536 mutants (Leu, Asn or Gln) resulted in an accumulation of the branched intermediate (Figure 1C) which failed to proceed to form spliced products, but instead decayed to form N-terminal cleavage products at a higher temperature or upon longer incubation (Figure 1C and Table II). Mutation of His536 also inhibited C-terminal cleavage. The branched intermediate produced by the His536Leu mutation was interconvertible with the linear precursor in a pH-dependent manner similar to that reported for wild-type MIP (Xu *et al.*, 1993), albeit at a slower rate and with more N-terminal cleavage (our unpublished data). The branched intermediate from the His536Leu mutant also had the same alkali-labile linkage between M and IP observed in the wild-type branched intermediate (Xu *et al.*, 1994).

Mutation of Asn537 to Asp (Figure 1B), Gln or Ala prevented splicing and cleavage at the intein C-terminus, but allowed cleavage at the intein N-terminus (Figure 1B and Table I). The branched intermediate was detected in both the Asn537Asp and Asn537Gln mutants.

Mutation of Ser538 by Cys, Thr, Ala and Gly predominantly affected branch formation and resulted in cleavage at both splice junctions (Tables I and II). Ser538Cys resulted in only trace amounts of ligated product (MP) after incubation for 16 h at 37°C or 50°C. Initially, cleavage at the N-terminal splice junction was more prevalent than cleavage at the C-terminal splice junction. With time, these single junction cleavage products were further cleaved to yield M + I + P. For example, the Ser538Ala precursor was cleaved initially at either splice junction, but predominantly yielded M + IP and to a lesser extent, MI + P ($t_{1/2}$ of ~4 h at 37°C or 0.5 h at 65°C); with longer incubations (16 h at 37°C or 4 h at 65°C), MI and IP were cleaved further to yield three segments, M, I and P (Table I).

Mutation of multiple residues at the C-terminal splice junction which prevented branch formation or C-terminal cleavage, did not prevent N-terminal cleavage (Tables I and II). The MIP precursor from a His536Leu/Asn537Ala double mutant and from two triple mutants (Ser1Cys/His536Leu/Ser538Ala and His536Leu/Asn537Ala/Ser538Ala) all yielded N-terminal cleavage products. When an Ochre stop codon was introduced at Ser538 or Gly539, incubation of the truncated MBP-intein fusion protein, MI, resulted in cleavage at the N-terminal splice junction when Ser or Cys were present at the intein N-terminus (Figure 2). However, substitution of Ser1 in MI by Ala blocked cleavage (Table III). Finally, a Ser1Ala and Asn537Ala MIP double mutant yielded a stable precursor which neither spliced, formed a branched intermediate nor cleaved at either junction (Table I).

Mutation of Ile2, Leu533, Tyr534 and Ala535

A Leu533Met substitution had little, if any, effect on splicing (Table I). However, substitution of the adjacent Tyr534 by Met resulted in a mutant that spliced normally, but was cold-sensitive (Figure 1D). The Tyr534Met mutant failed to splice at 30°C, but spliced efficiently at 42°C (Table II). Mutations which introduced a positively-charged lysine near either splice junction (Ile2Lys or Ala535Lys) resulted in cleavage at the C-terminal splice junction (Table I). Approximately 50% of the Ile2Lys precursor was cleaved *in vivo* after an overnight induction at 12–15°C. However, combination of the cold-sensitive Tyr534Met mutation with the Ile2Lys mutation resulted in a more stable MIP precursor after a similar overnight induction at 12–15°C; incubation of the Tyr534Met/Ile2 Lys purified precursor *in vitro* at 37°C resulted in production of MI + P with a $t_{1/2}$ of 4 h.

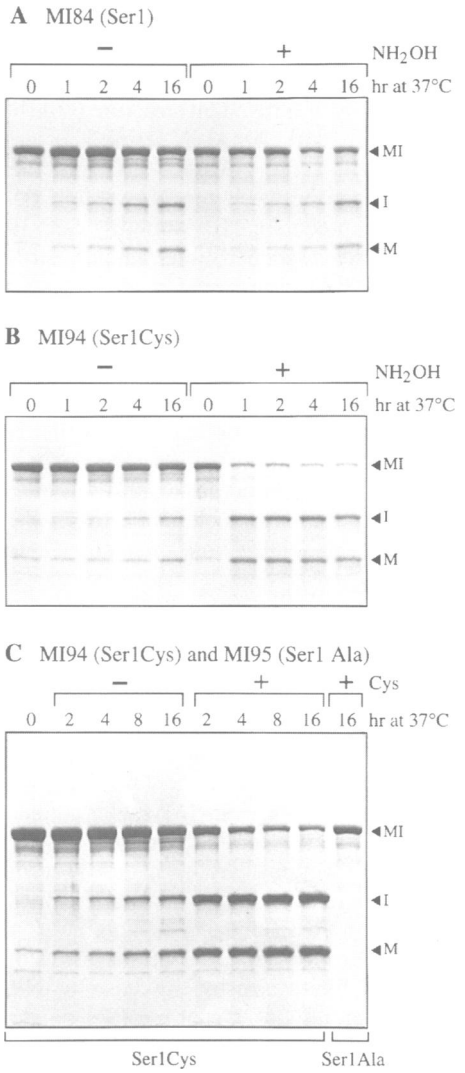

Fig. 2. Cleavage at the N-terminal splice junction by hydroxylamine and cysteine. (**A** and **B**) Hydroxylamine treatment. Truncated MI precursors with either Ser (A) or Cys (B) at the N-terminal of the intein were incubated at 37°C in the presence (+) or absence (−) of 0.25 M hydroxylamine (NH_2OH) at pH 6. (**C**) Cleavage of MI precursors with cysteine. MI94 (Ser1Cys) and MI95 (Ser1Ala) were treated with 50 mM cysteine at pH 7 for 0 to 16 h at 37°C to induce cleavage at the intein N-terminus.

Identification of a linear ester intermediate

The requirement of a hydroxyl- or thiol-containing residue (Ser, Cys or Thr) at the N-terminus of the intein for efficient splicing or N-terminal cleavage suggested that activation of the N-terminal splice junction may involve acyl rearrangement resulting in ester formation. Direct evidence of a linear ester intermediate was provided by experiments employing nucleophilic reagents. These experiments were based on the premise that an ester or thioester bond can be cleaved by treatment with weak nucleophiles which do not normally cleave peptide bonds and that thioester bonds are more susceptible to hydroxylamine cleavage than oxygen esters at a neutral or slightly acidic pH (Jencks *et al.*, 1960; Bruice and Benkovic, 1966; Bolanowski *et al.*, 1984). Truncated MI precursors that ended after Asn537 (Ser538Stop), and had either Ser, Ala or Cys at position 1, were treated with 0.25 M hydroxylamine (Figure 2). In the absence of hydroxylamine, the rates of N-terminal hydrolysis at pH 6 in MI84 (Ser1) and MI94 (Ser1Cys) were quite similar (Table III). In the presence of hydroxylamine, the rate of cleavage of MI84 (Ser1) increased only ~1.5-fold, the rate of cleavage of MI94 (Ser1Cys) increased ~112-fold and the Ser1Ala MI mutant was not cleaved at all (Table III). When Ser538 was included at the end of the truncated MI precursors, the rate of hydrolysis in MI83 (Ser1) and MI117 (Ser1Cys) was ~10-fold higher than in the respective constructs which lacked Ser538 (Table III). The rate of hydroxylamine induced N-terminal cleavage in the MI Ser1Cys mutants was the same in the presence or absence of Ser538.

Table III. Effect of hydroxylamine on the rate of N-terminal cleavage

Construct	Mutation	$t_{1/2}$ (h)		Increased rate[a]
		−	+	
MI84	Ser538Stop	41	27	1.5
MI94	Ser1Cys/Ser538Stop	37	0.33	112.1
MI95	Ser1Ala/Ser538Stop	–[b]	–[b]	
MI83	Gly539Stop	3.9	2.9	1.3
MI117	Ser1Cys/Gly539Stop	3.8	0.31	12.3

Truncated MI fusions containing an Ochre stop codon at the indicated positions were treated with 0.25 M hydroxylamine (HA) at 37°C, pH 6 and the production of N-terminal cleavage products with time was determined after scanning Coomassie Blue stained gels. The $t_{1/2}$ for N-terminal splice junction cleavage is listed in the presence (+) or absence (−) of hydroxylamine.
[a]Increase in $t_{1/2}$ values after hydroxylamine treatment [(HA −sample) divided by (HA + sample)].
[b]Cleavage products were not observed.

The N-terminal splice junction was also sensitive to cleavage by the free amino acid, cysteine. Cys added in solution was able to accelerate the rate of N-terminal cleavage in MI94 (Ser1Cys) by ~20-fold at pH 7 (Figure 2), but had no significant effect on the rate of N-terminal cleavage in MI84 (Ser1) (our unpublished data). MI95 (Ser1Ala) precursor was stable in the presence of Cys (Figure 2). Treatment with Ser or Ala in solution had no affect on cleavage of any MI precursor, indicating that the sulfhydryl side chain of exogenous cysteine is responsible for cleavage, rather than the α-amino group. Finally, incubation in 30 mM dithiothreitol resulted in N-terminal cleavage of MI94 (our unpublished data).

Proposed mechanism of protein splicing and cleavage side reactions

The data presented in this and previous studies support the protein splicing mechanism depicted in Figure 3 (pathway B in Xu *et al.*, 1994). Although the splicing of MIP containing the Psp pol intein-1 with Ser at both splice junctions is depicted, the mechanism is equally applicable to other inteins containing Thr or Cys at these positions. However, differences in the chemical properties of these three amino acids may affect the relative rates of each step or the intermediates observed, including the branched intermediate. For example, the sulfhydryl group of Cys is a much better leaving group than the hydroxyl group of Ser when the incoming nucleophile is an amino or sulfhydryl group and thus, thioesters are more labile than oxygen esters under these conditions.

The first step in the protein splicing pathway is activation

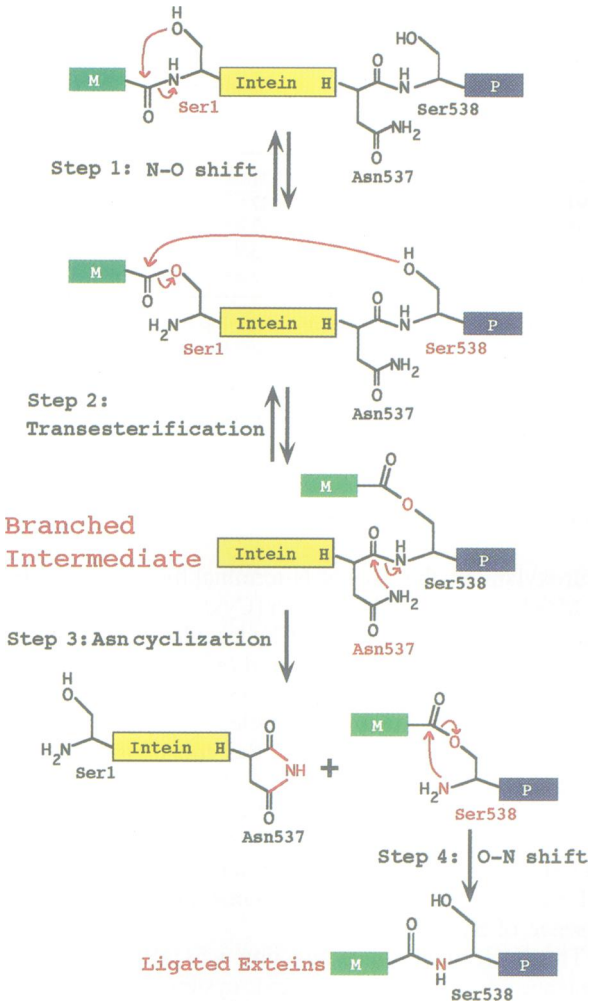

Fig. 3. Proposed mechanism of protein splicing. This mechanism was originally proposed in Xu *et al.* (1994) as pathway B. Although this figure depicts the splicing of MIP containing the Psp pol intein-1 (537 amino acids) with Ser at both splice junctions, it would be equally applicable to other inteins containing Thr or Cys at these positions. All proton transfer steps and tetrahedral intermediates have been omitted for clarity. The first step in the protein splicing pathway is an N–O acyl shift at Ser1. This acyl rearrangement moves the N-extein (M) to the side chain of Ser1. Next, transesterification occurs when the upstream ester bond is attacked by the hydroxyl group of Ser538 resulting in transfer of the N-extein (M) to the side chain of Ser538, forming the branched intermediate. The branch is resolved by cyclization of the conserved intein C-terminal Asn to form a succinimide. His536 assists in Asn cyclization. Succinimide formation leaves the N-extein (M) attached to the side chain of Ser538 via an ester bond which is converted to a normal peptide bond by a spontaneous O–N acyl shift. Part or all of the remaining residues of the intein are required for proper folding of the intein to generate the active site. Other intein residues or water molecules assist these chemical reactions. (See text for a detailed explanation of this mechanism.) The green box represents the N-extein or M, in the case of MIP. The blue box represents the C-extein or P, in the case of MIP. The H in the yellow intein box indicates the position of the conserved His536.

of the N-terminal splice junction. An N–O acyl rearrangement at Ser1 moves the N-extein (M) to the side chain of Ser1, forming a linear ester intermediate. Mutation data presented above indicate that ester formation at the N-terminal splice junction does not require any of the conserved residues except Ser1, although it may be facilitated by Ser538 (Table III). Next, transesterification occurs when the upstream ester bond is attacked by the hydroxyl group of Ser538 resulting in cleavage at the N-terminal splice junction and transfer of the N-extein (M) to the side chain of Ser538. The resultant structure is the branched protein intermediate with: (i) two N-termini corresponding to the N-terminus of the N-extein (M) and the intein, (ii) an ester linkage between the N-extein (M) and the remainder of the precursor (IP) and (iii) Ser538 as the branch point. This branched intermediate structure is consistent with the previously reported properties of the branched intermediate (Xu *et al.*, 1993, 1994). Branch formation requires the presence of an ester at the N-terminal splice junction and Ser 538, but not Asn537 nor His536 (Figure 1). Therefore, mutations which prevent the acyl rearrangement at the upstream splice junction will also prevent branch formation. Depending on the intein and its context, conservative substitution (Ser, Thr or Cys) of the downstream nucleophile (Ser538 in MIP) may also yield a branched intermediate (Table I and Hodges *et al.*, 1992). The branch is resolved by cyclization of the conserved intein C-terminal Asn to form a succinimide ring. Asn cyclization cleaves the peptide bond between the intein and the C-extein (P). His536 assists in Asn cyclization and branch resolution (Figure 1). Release of the intein by Asn cyclization leaves the N-extein (M) attached to the side chain of Ser538 via an ester bond which is resolved by a spontaneous O–N acyl rearrangement resulting in the N-extein (M) being linked to the C-extein (P) via a native peptide bond.

Possible side reactions are isolated N-terminal or C-terminal cleavage (Figure 4) and hydrolysis of the succinimide to form Asn and iso-asparagine. C-terminal cleavage results from Asn cyclization prior to branch formation and is likely assisted by His536. N-terminal cleavage results from hydrolysis of the ester bond in either the linear ester intermediate or the branched intermediate.

Part or all of the remaining residues of the intein are required for proper folding of the intein to generate the active site. Other intein residues or water molecules probably assist these chemical reactions by stabilizing intermediates and/or accepting a proton from the nucleophilic attacking group or donating a proton to the leaving group either directly or via a charge relay system.

Discussion

To understand the basis of protein splicing, we have studied the chemical reactions involved and the sequence of these reactions by isolating, identifying and analyzing splicing precursors, intermediates and products. We have examined closely the protein splicing pathway after mutagenesis of eight residues proximal to the splice junctions and have assigned roles to all of the conserved junction residues.

The mutagenesis and chemical cleavage data provide strong support for an N–O (N–S) acyl rearrangement as the activation mechanism at the N-terminal splice junction. Substitution of Ser1 with amino acids which are unable to undergo acyl rearrangement blocked N-terminal cleavage and branch formation. N-terminal cleavage in the absence of branch formation (Ser538Ala, Ser538Gly, His536Leu/Asn537Ala, His536Leu/Asn537Ala/Ser538-Ala, MI84 and MI94 mutants) also suggests the presence

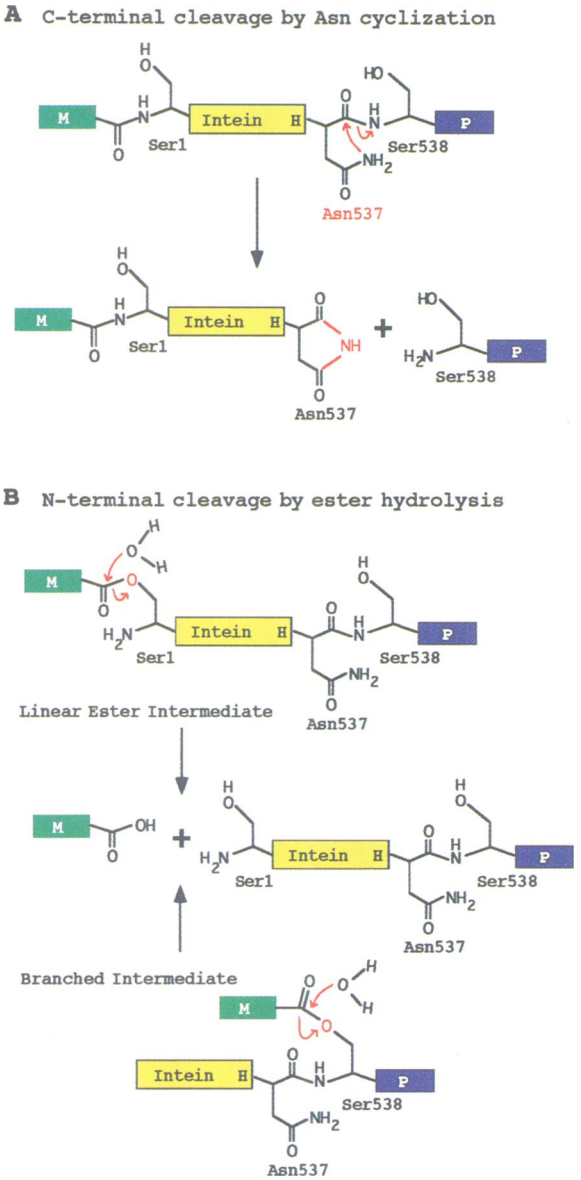

Fig. 4. Proposed N-terminal and C-terminal cleavage side reactions. (**A**) C-terminal cleavage results from Asn cyclization prior to branch formation. Asn cyclization can also occur prior to or after N-terminal cleavage. (**B**) N-terminal cleavage results from hydrolysis of the ester bond in the linear ester or branched intermediates. The green box represents the N-extein or M, in the case of MIP. The blue box represents the C-extein or P, in the case of MIP. The H in the yellow intein box indicates the position of the conserved His536.

of a linear ester intermediate (Figure 4). Moreover, none of the conserved C-terminal residues were necessary for N-terminal cleavage (His536Leu/Asn537Ala/Ser538Ala mutant and Table I). Taken together, these mutagenesis data indicate that Ser1 is essential for initiation of the splicing pathway and N-terminal cleavage. However, MI constructs including Ser538 were hydrolyzed at faster rates than constructs lacking Ser538. This suggests that either Ser538 assists in formation, stabilization or hydrolysis of the ester at Ser1 or that the increased rate of N-terminal cleavage is due to hydrolysis of the potential branched intermediate formed in the presence of Ser538. Introduction of a positively-charged Lys near either splice junction (Ile2Lys or Ala535Lys) prevented cleavage at the N-terminal splice junction, possibly by interfering with ester formation or attack by the downstream nucleophile. Inhibition of N-terminal cleavage due to the presence of a Lys at A535 suggests further that its side chain is positioned very close to Ser1 in the folded intein.

More direct evidence of the linear ester intermediate was provided by experiments using nucleophilic reagents to cleave the ester bond. Hydroxylamine strongly stimulated N-terminal cleavage in the Ser1Cys mutant (~112-fold increase), weakly stimulated cleavage in the Ser1 precursor, but failed to cleave the Ser1Ala mutant. These data are indicative of the presence of an ester/thioester due to an acyl shift at Ser1 (or Ser1Cys), since hydroxylamine cleaves thioesters at a much faster rate than oxygen esters and does not cleave peptide bonds under these reaction conditions (Jencks *et al.*, 1960; Bruice and Benkovic, 1966; Bolanowski *et al.*, 1984). Detection of a hydroxamate at the C-terminus of M by high resolution mass spectrometry after hydroxylamine treatment of MIP confirmed that the hydroxylamine-sensitive site is the bond at the N-terminal splice junction (Shao *et al.*, 1996). N-terminal splice junction cleavage experiments using free amino acids in solution indicated that: (i) the free Cys is attacking a thioester bond in the Ser1Cys precursors since the sulfhydryl group of Cys is unlikely to cleave a peptide bond and (ii) that the thiol group of cysteine, not its amino group, is the attacking nucleophile since free Ala or Ser did not cleave MI94 (Ser1Cys). The demonstration that the sulfhydryl group of Cys can cleave the upstream thioester bond supports the transesterification mechanism. Ser in solution probably failed to cleave the thioester bond in MI94 because, due to its pK_a of ~13.7, Ser hydroxyl groups would not have been deprotonated at pH 7 and thus, would have been unable to act as nucleophiles, whereas the thiol group of Cys, with a pK_a of 8.3, can act as a nucleophile at pH 7. However, the deprotonation of Ser1 or Ser538 in the context of the intein active site is determined by the local environment of that residue and not by the pH of the solution. The pK_a difference between Cys and Ser may also explain why these structurally conservative substitutions drastically affect splicing. Inteins naturally containing Ser would most probably require assistance from surrounding residues to deprotonate the attacking nucleophile since, owing to its high pK_a, the hydroxyl group on the side chain of Ser would not be significantly deprotonated at the pH optimum of the splicing reaction. Therefore, inteins naturally containing serines at either splice junction need to provide a network to enhance the deprotonation of the serine hydroxyl group. Cysteine, on the other hand, might not need any special local environment to act as a nucleophile because of its lower pK_a. It should be noted that substitution of Cys with Ser never yields spliced products (Davis *et al.*, 1992; Hirata and Anraku, 1992; Cooper *et al.*, 1993) whereas substitution of Ser or Thr with Cys may yield varying amounts of spliced products (Table I and Hodges *et al.*, 1992).

The observation that N-terminal cleavage, branch formation and splicing have very similar pH profiles suggests that the pH optimum of the overall splicing reaction is governed predominantly by the pH requirement of this first step. A similar acyl shift activation mechanism of peptide bond cleavage has been described in the autocleavage of prohistidine decarboxylase (Vanderslice

et al., 1988; Van Poelje and Snell, 1990), glycosylasparaginase (Guan *et al.*, 1996) and the hedgehog protein (Porter *et al.*, 1995). A similar reduction in the rate of cleavage was observed in histidine decarboxylase when Ser was replaced by Cys (8-fold slower) or Thr (450-fold slower) (Vanderslice *et al.*, 1988).

After formation of the ester intermediate at the N-terminal splice junction, Ser538 is the nucleophile responsible for branch formation. We have previously shown that the branched intermediate is resolved by cyclization of the intein C-terminal Asn (Xu *et al.*, 1994). Mutagenesis data indicate that the penultimate His is required to assist branch resolution (Figure 1C), most probably by acting as a proton donor and/or acceptor. Branch resolution leaves the N-extein linked to the side chain of Ser538. We have proposed that a native peptide bond is formed by spontaneous acyl rearrangement. Again, there are several lines of evidence in the literature which support this type of spontaneous acyl shift required to re-establish the native peptide bond between the two exteins. Peptide studies have shown that model peptides can spontaneously undergo an O–N acyl shift (Xu *et al.*, 1994). A similar spontaneous acyl rearrangement to form a native peptide bond has been indicated in the 'native chemical ligation' method (Dawson *et al.*, 1994).

The data presented are incompatible with the Cooper and Stevens mechanism (Cooper *et al.*, 1993), the Clarke mechanism (Clarke, 1994) and the branch formation pathway A mechanism (Xu *et al.*, 1994). In the absence of branch formation, none of these models include a linear ester intermediate nor provide a mechanism for N-terminal cleavage unless one invokes an environmentally-induced general peptide bond lability. The Cys cleavage data disfavor any such peptide bond lability since the thiol group of Cys is unlikely to cleave a peptide bond. The structure of the branched intermediate also precludes the Cooper and Stevens mechanism (Cooper *et al.*, 1993). Detection of a stable branched intermediate in Asn537Asp samples after electrophoresis precludes the Clarke mechanism (Clarke, 1994) unless one invokes a novel mechanism for stabilizing the resultant anhydride. The data are also incompatible with the early models (Wallace, 1993; Xu *et al.*, 1993) which suggest that splicing is initiated by an N–O acyl shift at both splice junctions, followed by attack on the N-terminal ester by the amino group of Ser538 (in the acyl shifted ester form). On the other hand, all the data are consistent with the transesterification mechanism (Figure 3). Furthermore, the transesterification mechanism has support from experiments in protein semisynthesis. The native chemical ligation mechanism is completely analogous to the mechanism proposed in Figure 3 with the exception that the two 'peptides' are joined by the intein which is removed by the cyclization of the conserved Asn (Dawson *et al.*, 1994).

The data presented define the roles of the four conserved splice junction residues. The role of Ser1 is to initiate splicing by undergoing an acyl rearrangement. The role of the side chain of Ser538 is to act as a nucleophile to cleave the ester at Ser1 and form the branched intermediate. The role of His536 is to assist in the cyclization of Asn537. The role of Asn537 is branch resolution. Similar mutations in other inteins have produced similar effects, except for the inability to detect branched intermediates or to splice *in vitro* (Davis *et al.*, 1992; Hirata and Anraku, 1992; Hodges *et al.*, 1992; Cooper *et al.*, 1993). For example, mutation of Cys1 to Ser in the Sce VMA intein blocked N-terminal cleavage and resulted in C-terminal cleavage (Hirata and Anraku, 1992; Cooper *et al.*, 1993). In the Tli pol intein-2 native system, substitution of Ser1 by Thr resulted in significant splicing, substitution of Ser1 by Cys yielded small amounts of spliced product and substitution of Ser1 by Ala resulted in C-terminal cleavage (Hodges *et al.*, 1992). Mutation of the C-terminal Cys in both the Sce VMA intein and the Mtu recA intein reduced (Cys to Ser) or prevented (Cys to Gly, Val or Arg) splicing (Davis *et al.*, 1992; Hirata and Anraku, 1992; Cooper *et al.*, 1993). In the Sce VMA intein, mutation of the penultimate His also prevented or reduced splicing and mutation of the conserved Asn blocked splicing (Cooper *et al.*, 1993).

Self-splicing suggests that the reaction is an isoenergetic process requiring efficient coupling of bond cleavage and formation. The sensitivity of the splicing reaction to subtle differences among the nucleophiles at each splice junction (Ser, Cys or Thr) indicates that these positions have a very strict chemical, electronic and steric specificity. The experiments presented here have increased our understanding of the chemical mechanism of protein splicing, but some parameters of the reaction still need to be worked out. It is assumed, for example, that residues throughout the intein assist the chemical reactions by stabilizing intermediates or acting as proton donors or acceptors. The conserved His in Block B (Pietrokovski, 1994) is a candidate for these functions.

These experiments also suggest potential uses of protein splicing in protein engineering and protein purification. The cold-sensitive mutant (Tyr534Met) is a good starting point for developing a conditional knockout system based on temperature controlled splicing *in vivo*. Mutation of the intein N-terminus (Ser1 in MIP) should convert any intein into a C-terminal cleavage element. Mutation of the intein C-terminal Asn and/or penultimate His should convert any intein into a N-terminal cleavage element. If cleavage is controllable, these elements could be utilized for protein purification when combined with an affinity tag and a multiple cloning site for the target protein. After affinity purification, the protein of interest could then be cleaved from the remainder of the fusion without requiring addition of exogenous proteases. Understanding the mechanism of protein splicing may allow the future development of elements which cleave or ligate proteins at will.

Materials and methods

Numbering of residues in MIP

Amino acid numbers refer to the position in the Psp pol intein with Ser1 being the first residue of the intein and Asn537 the last. The numbering continues sequentially into the C-extein (P) beginning with Ser538. This convention was adopted to allow easy comparison of intein residues in different protein contexts.

Mutagenesis, protein purification and in vitro splicing

The splice junction mutants were derived from pMIP-21 by cassette replacement of splice junction sequences as previously described (Xu *et al.*, 1994). pMIP-21 contains a *Xho*I and a *Kpn*I site flanking the N-terminal splice junction and a *Bam*HI and *Stu*I site flanking the C-terminal splice junction. Briefly, cassette replacement involved

digesting pMIP-21 with the appropriate restriction enzymes for each splice junction, and then replacing the deleted region with double-stranded oligomers containing the desired mutation. MIP purification and splicing conditions have been previously described (Xu *et al.*, 1993, 1994). Splicing of MIP can be controlled by temperature and, *in vitro*, by pH. MIP mutants were induced at low temperature (12–15°C). After induction, MIP precursor was purified at pH 8.5 in a single step by affinity chromatography on amylose resin (New England Biolabs). Splicing of MIP was stimulated by incubation in 20 mM sodium phosphate pH 6, 0.5 M NaCl for the designated times and temperatures (30–65°C).

Hydroxylamine, cysteine and dithiothreitol treatment

Truncated MI proteins terminating in Asn537 or Ser538 were purified over amylose resin in pH 6 column buffer (20 mM sodium phosphate pH 6, 0.5 M NaCl). Samples were incubated at 37°C for up to 16 h in the presence or absence of 0.25 M hydroxylamine. Samples were adjusted to pH 6 with 0.1 M Bis-Tris-Propane after addition of hydroxylamine. As little as 10 mM hydroxylamine enhanced significantly the cleavage of MI94. Cysteine cleavage reactions with truncated MI proteins were carried out at 37°C for up to 16 h in 20 mM Na–HEPES pH 7, 0.5 M NaCl, 50 mM cysteine. The presence of 30 mM dithiothreitol also increased the rate of cleavage of MI94 at 37°C.

Electrophoresis and quantitation

Samples were electrophoresed on 4–12% SDS–polyacrylamide gels (Novex) and the gels were stained with Coomassie Blue. Protein Marker, Broad Range (New England Biolabs) standards were used. Coomassie Blue stained gels were digitized with a Microtek Scanmaker III and analyzed with NIH Image 1.51 software for quantitation and rate determinations.

Acknowledgements

We thank Donald Comb for support and encouragement, Melissa Scott, Fana Mersha and Maurice W.Southworth for technical assistance, and Donald Comb, Sharong Chong, Virginia W.Cornish, Fana Mersha, Chris Noren, Henry Paulus, Peter Schultz, Yang Shao and Maurice W.Southworth for valuable discussions.

References

Bolanowski,M.A., Earles,B.J. and Lennarz,W.J. (1984) Fatty acylation of proteins during development of sea urchin embryos. *J. Biol. Chem.*, **259**, 4934–4940.

Bruice,T.C. and Benkovic,S.J. (1966) *Bioorganic Mechanisms*. W.A.Benjamin, New York, Vol. I, pp. 268–294.

Clarke,N.D. (1994) A proposed mechanism for the self-splicing of proteins. *Proc. Natl Acad. Sci. USA*, **91**, 11084–11088.

Cooper,A.A., Chen,Y., Lindorfer,M.A. and Stevens,T.H. (1993) Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J.*, **12**, 2575–2583.

Davis,E.O., Sedgwick,S.G. and Colston,M.J. (1991) Novel structure of the recA locus of *Mycobacterium tuberculosis* implies processing of the gene product. *J. Bacteriol.*, **173**, 5653–5662.

Davis,E.O., Jenner,P.J., Brooks,P.C., Colston,M.J. and Sedgwick,S.G. (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.

Davis,E.O., Thangaraj,J.S., Brooks,P.C. and Colston,M.J. (1994) Evidence of selection for protein introns in the RecAs of pathogenic mycobacteria. *EMBO J.*, **13**, 699–703.

Dawson,P.E., Muir,T.W., Clark-Lewis,I. and Kent,S.B. (1994) Synthesis of proteins by native chemical ligation. *Science*, **266**, 776–779.

Gimble,F.S. and Thorner,J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, **357**, 301–306.

Gu,H.H., Xu,J., Gallagher,M. and Dean,G.E. (1993) Peptide splicing in the vacuolar ATPase subunit A from *Candida tropicalis*. *J. Biol. Chem.*, **268**, 7372–7381.

Guan,C., Cui,T., Rao,V., Liao,W., Benner,J., Lin,C.L. and Comb,D. (1996) Activation of glycosylasparaginase. Formation of active N-terminal threonine by intramolecular autoproteolysis. *J. Biol. Chem.*, **271**, 1732–1737.

Hirata,R. and Anraku,Y. (1992) Mutations at the putative junction sites of the yeast VMA1 protein, the catalytic subunit of the vacuolar membrane H^+-ATPase, inhibit its processing by protein splicing. *Biochem. Biophys. Res. Commun.*, **188**, 40–47.

Hirata,R., Ohsumi,Y., Nakano,A., Kawasaki,H., Suzuki,K. and 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.

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.

Jencks,W.P., Cordes,S. and Carriulos,J. (1960) The free energy of thiol ester hydrolysis. *J. Biol. Chem.*, **235**, 3608–2314.

Kane,P.M., Yamashiro,C.T., Wolczyk,D.F., Neff,N., Goebl,M. and Stevens,T.H. (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase. *Science*, **250**, 651–657.

Lambowitz,A.M. and Belfort,M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem.*, **62**, 587–622.

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.

Pietrokovski,S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci.*, **3735**, 2340–2350.

Porter,J.A., Kessler,D.P., Ekker,S.C., Young,K.E., Lee,J.J., Moses,K. and Beachy,P.A. (1995) The product of hedgehog autoproteolytic cleavage active in local and long-range signalling. *Nature*, **374**, 363–366.

Shao,Y., Xu,M. 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. 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.

Van Poelje,P.D. and Snell,E.E. (1990) Cloning, sequencing, expression, and site-directed mutagenesis of the gene from *Clostridium perfringens* encoding pyruvoyl-dependent histidine decarboxylase. *Biochemistry*, **29**, 132–139.

Vanderslice,P., Copeland,W.C. and Robertus,J.D. (1988) Site-directed alteration of serine 82 causes nonproductive chain cleavage in prohistidine decarboxylase. *J. Biol. Chem.*, **263**, 10583–10586.

Wallace,C.J. (1993) The curious case of protein splicing: Mechanistic insights suggested by protein semisynthesis. *Protein Sci.*, **2**, 697–705.

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 on March 20, 1996; revised on June 17, 1996