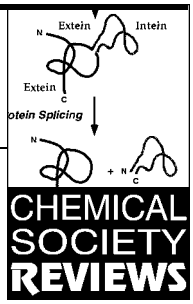
The chemical basis of protein splicing



Henry Paulus

Boston Biomedical Research Institute, Boston, MA 02114 and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

Protein splicing is a recently discovered mechanism for the post-translational processing of proteins. It involves the self-catalyzed excision of an intervening polypeptide, the intein, from an inactive enzyme precursor and the formation of an active enzyme by joining the flanking regions by a peptide bond. Protein splicing occurs at a catalytic center that resides entirely within the intein. The catalyzed reactions include rearrangement of a peptide bond adjacent to cysteine or serine to yield a peptide ester, intramolecular transesterification involving a second cysteine, serine, or threonine side chain to yield a branched protein, and cyclization of an asparagine residue coupled to peptide bond cleavage to effect intein excision. This review discusses the mechanisms of these reactions and of similar reactions that underlie other types of protein rearrangements as well as the current state of knowledge on how these reactions are catalyzed.

1 Introduction

Since the discovery of zymogens in the 1930s, it has been clear that many proteins are synthesized as inactive precursors that are activated by the rearrangement of peptide bonds. Such post-translational processing is usually catalyzed by other proteins, as in the activation of the blood clotting factors by highly selective proteases. However, in the last 10 years, many examples of self-catalyzed peptide bond rearrangements have been described. These include protein splicing, the autoprocessing of hedgehog proteins, the autocleavage of amidohydrolase precursors, and the formation of pyruvoyl enzymes. A common feature of these self-catalyzed reactions is that they are initiated by the N→S or N→O acyl rearrangement of a peptide bond involving the amino group of cysteine, serine, or threonine. Moreover, protein splicing elements and the autoprocessing domains of hedgehog proteins have significant structural and sequence homology, suggesting an evolutionary relationship.[1] Protein splicing, which occurs in many types of organisms and whose mechanism is now relatively well understood, will thus serve as the paradigm for self-catalyzed protein rearrangements. As we discuss the mechanism of protein splicing, we will point out similarities with the reactions that underlie other types of protein rearrangements.

Protein splicing was discovered in 1990, first in yeast, a unicellular eukaryote, but then also in bacteria and archaea. About 50 examples of protein splicing are known to date, of which 36 are listed in the latest published compilation.[2] Protein splicing involves the excision of an intervening polypeptide segment (the 'intein'), which usually interrupts a functional domain of an enzyme, and the joining of the flanking regions (the 'exteins') through a peptide bond, thereby generating an active enzyme. This process is formally analogous to RNA splicing, in which an intervening sequence (the 'intron') is excised and the flanking regions (the 'exons') are joined by a normal 3′,5′-phosphodiester bond to yield a functional messenger RNA molecule (see Fig. 1).

RNA splicing and protein splicing may be viewed as alternative methods for the expression of interrupted genes, which differ only in the level at which excision of the intervening sequences occurs. From this perspective, both introns and inteins are manifestations of 'selfish DNA', whose excision is essential for the survival of the organism. Many introns and inteins also harbor homing endonucleases, which can initiate the insertion of the DNA sequences encoding these elements into other genes either in the same or in different organisms. Such intervening sequences can be considered parasites that insert themselves into vital host genes yet avoid killing their hosts because they can excise themselves and thereby restore host gene function.

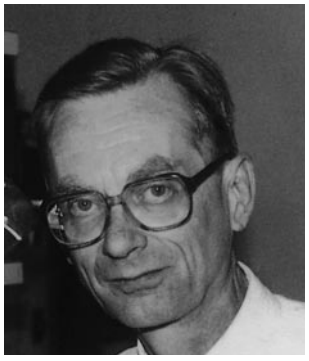
RNA splicing and protein splicing may also have regulatory functions. Introns are generally positioned in regions of genes that correspond to domain boundaries in the encoded proteins, and alternate RNA splicing provides a mechanism for the modular assembly of functionally different proteins from a single gene. In contrast, inteins are usually embedded in highly conserved regions of functional protein domains and protein splicing is a mechanism for the activation of an inactive precursor protein. Both types of intervening sequences can thus serve to control gene expression, albeit to different ends. Because there are introns and inteins that lack homing endonuclease domains, one might speculate that the primary function of intervening sequences is indeed regulatory but that they are frequently harnessed by homing endonucleases as vehicles for horizontal transmission to other organisms and species. Perhaps the strongest argument for an origin of inteins as regulatory elements is the homology of the protein splicing domains to the autoprocessing domains of hedgehog proteins. The latter occur in essentially all metazoans, ranging from nematodes and fruit flies to man, and fulfill a purely regulatory function in early developmental pattern formation. This suggests that protein splicing elements, known to occur only in unicellular organisms, were modified early in the evolution of multicellular organisms to regulate cell fate in embryonic development.

2 Chemical strategy of protein splicing

2.1 General characteristics of the protein splicing process

The investigation of the chemical mechanisms underlying protein splicing was made possible by the classic experiments

Henry Paulus was educated at the University of Chicago, where he carried out research for the PhD (1959) under the guidance of Eugene P. Kennedy. He was an NSF postdoctoral fellow at Cambridge University with Alexander Todd and at Harvard University with Konrad Bloch. In 1961, he joined the Department of Biological Chemistry at the Harvard Medical School, where he still holds a faculty appointment, but in 1975 he moved his laboratory to the Boston Biomedical Research Institute where he is a Senior Scientist.



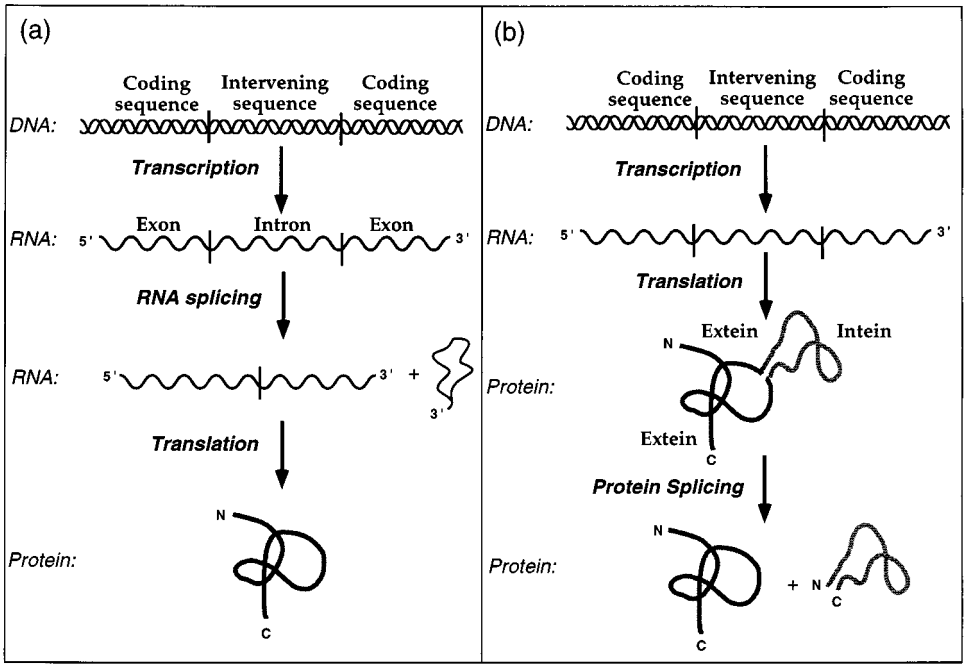


Fig. 1 Comparison of the excision of intervening sequences by (a) RNA splicing and (b) protein splicing

of Xu and coworkers.[3] The coding sequence for the intein embedded in the DNA polymerase of the hyperthermophilic archeon, *Pyrococcus* sp. GB-D, was inserted into a foreign context so as to allow synthesis of the unspliced precursor at low temperatures and the induction of splicing of the purified precursor by an increase in temperature. In these hybrid constructs, the maltose binding protein of *Escherichia coli* was used as the N-extein, to allow rapid affinity purification of the unspliced precursor on amylose resin. These experiments yielded two important insights into the general nature of protein splicing: (i) protein splicing occurs with highly purified precursor proteins in which the intein is inserted between exteins that bear no relationship to protein splicing and is expressed in an organism (*E. coli*) in which protein splicing normally does not occur, indicating that protein splicing requires no accessory enzymes and that all necessary catalytic groups are part of the intein itself; and (ii) protein splicing is induced merely by increasing the temperature in the absence of cofactors such as ATP, indicating that protein splicing is a spontaneous process requiring neither organic cofactors nor sources of metabolic energy.

Additional information on the nature of protein splicing came from the laboratory of Anraku,[4] who developed an *in vitro* protein splicing system based on the intein from the VMA subunit of the vacuolar ATPase of *Saccharomyces cerevisiae*. Expression of recombinant proteins containing this intein in *E. coli* leads to the accumulation of unspliced precursor in the form of inclusion bodies that can be solubilized with 6 M guanidinium chloride and undergo splicing upon removal of the denaturant. Using this experimental system, Anraku and coworkers made two important observations:[5] (i) protein splicing is refractory to inhibitors of cysteine-, serine-, aspartic-, and metallo-proteases, suggesting that the protein splicing catalytic center does not resemble the active sites of any of the four major classes of proteases, and (ii) no swapping of exteins occurs when mixtures of protein splicing precursors, in which the VMA intein was inserted between different pairs of exteins, are allowed to undergo splicing, indicating that protein splicing is exclusively an intramolecular reaction.

The general picture of protein splicing that emerges from these *in vitro* observations is as follows:

- protein splicing is catalyzed solely by amino acid residues contained in the intein;
- protein splicing requires no coenzymes or sources of metabolic energy and therefore involves bond rearrangements rather than bond cleavage followed by resynthesis;
- protein splicing is an intramolecular process;
- the amino acid residues that catalyze protein splicing are arranged in an active center which differs from that of any of the four major classes of proteases.

2.2 Summary of the reaction steps that underlie protein splicing

The general nature of these reactions was anticipated on the basis of the two types of amino acid side chains that are

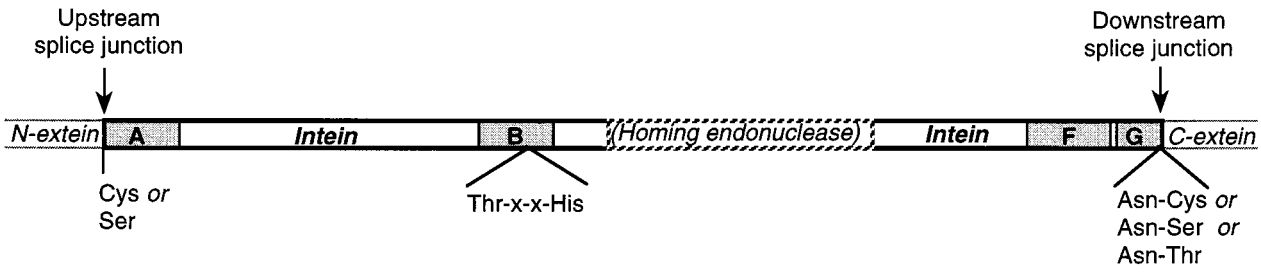


Fig. 2 Major sequence features of a typical protein splicing element inserted between an N-extein (shown partially on the left) and a C- extein (right). The bold rectangle represents the protein splicing element, with its N-terminus at the left, and the shaded boxes labeled A, B, F, and G, represent conserved sequence motifs common to all protein splicing elements. The size and positions of the rectangles are approximately to scale. Invariant amino acid residues that occur in all inteins are indicated below the rectangle. The portion of the intein enclosed in dotted lines represents the homing endonuclease domain, which is found in most but not all inteins; its size exceeds that of the protein splicing element considerably and is not represented on scale.

invariably found adjacent to the splicing junctions (see Fig. 2). (i) The N-termini of the intein and the C-extein are always amino acids that carry a nucleophilic side chain: cysteine (Cys), serine (Ser), or threonine (Thr). These amino acid side chains often function in nucleophilic attacks on amide and ester bonds, including—albeit usually under extreme conditions—attacks on amide bonds involving their own amino groups. (ii) At the C-terminus of the intein one always finds an asparagine (Asn) residue. In proteins with long biological half-lives or in synthetic peptides during the acidic conditions encountered in deprotection steps, Asn residues—especially when adjacent to glycine—often undergo spontaneous cyclization reactions that can lead to deamidation or peptide bond cleavage.

Detailed biochemical studies on protein splicing during the period 1993–96 succeeded in defining the chemical reactions that underlie this unusual catalytic process.[6–11] These reactions, which consist of the four steps summarized in Fig. 3, are described in the next four sections of this review. Briefly, these steps are:

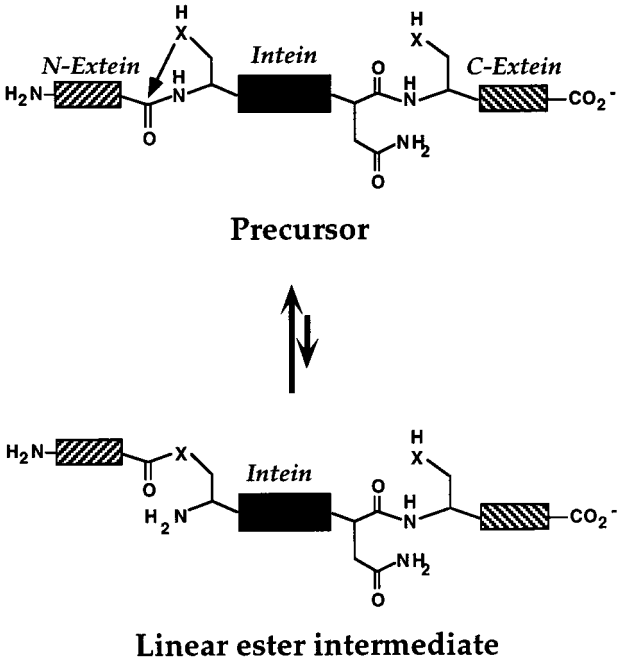
Step 1 Formation of a linear ester intermediate by N→S or N→O acyl rearrangement involving the nucleophilic amino acid residue at the upstream splice junction;

Step 2 Formation of a branched ester intermediate by the attack of the nucleophilic residue at the downstream splice junction on the linear ester intermediate;

Step 3 Cyclization of the asparagine residue adjacent to the downstream splice junction, coupled to cleavage of the branched ester intermediate to yield an excised intein with a C-terminal aminosuccinimide residue and the two exteins joined by an ester bond;

Step 4 Spontaneous hydrolysis of the aminosuccinimide residue and rearrangement of the ester linking the exteins to the more stable amide bond.

3 N→O and N→S acyl rearrangements (Scheme 1)



Scheme 1

3.1 Acid-induced N→O acyl rearrangements

When treated with strong acids at low temperatures, peptide bonds involving the amino groups of serine or threonine rearrange to ester bonds involving the corresponding amino acid side chains (reviewed in ref. 12). Such N→O acyl rearrangements occur through the nucleophilic attack by the amino acid side chain on the peptide carbonyl carbon and are thought to proceed through a hydroxyoxazolidine intermediate (Fig. 4). At

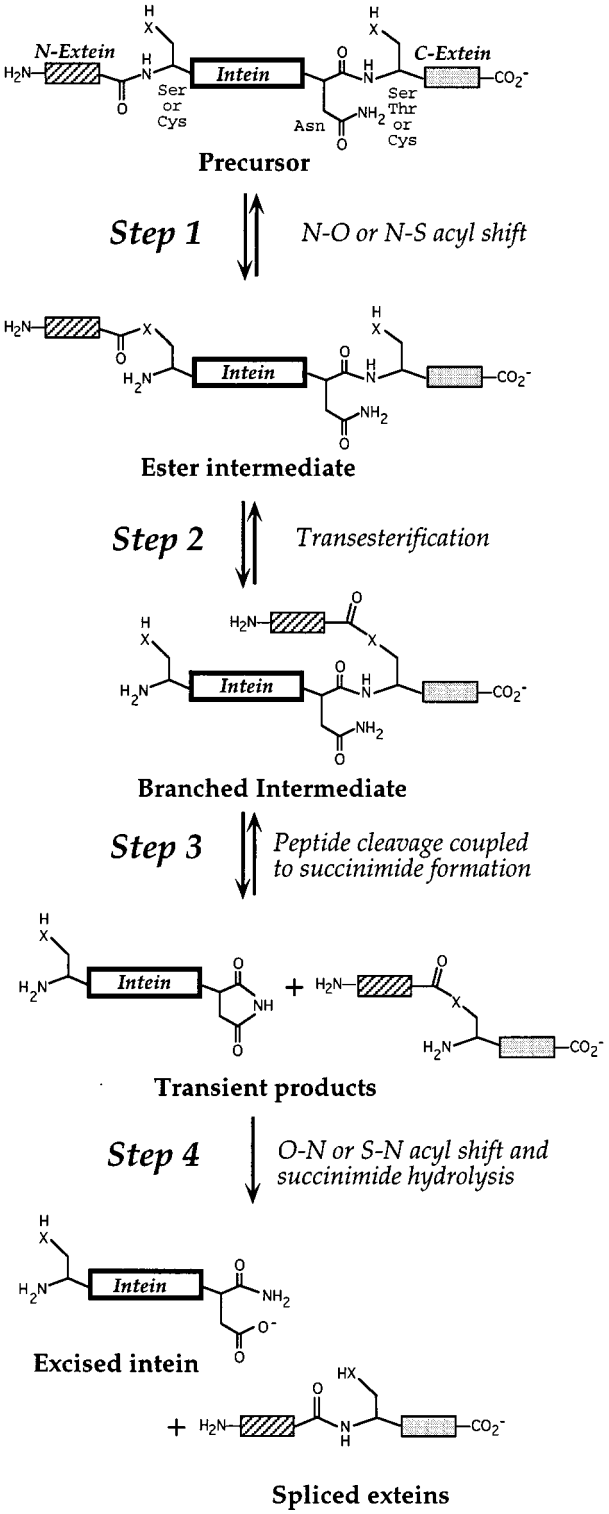


Fig. 3 Mechanism of protein splicing. The amino acid residues that participate directly in the chemical transformations are shown (X = O or S). The remainder of the intein and exteins are shown by boxes, which are not to scale.

neutral pH the equilibrium position for N→O acyl shifts favors the amide, but under strongly acidic conditions the ester is stabilized by protonation of the free amino group. Reversal of acid-induced N→O acyl shifts can be prevented by the acetylation, formylation, dinitrophenylation, or deamination of amino groups at low pH. This allows subsequent alkaline hydrolysis of the ester bonds, thereby achieving selective cleavage of proteins adjacent to Ser and Thr residues. Although N→S acyl rearrangements involving the analogous reaction of

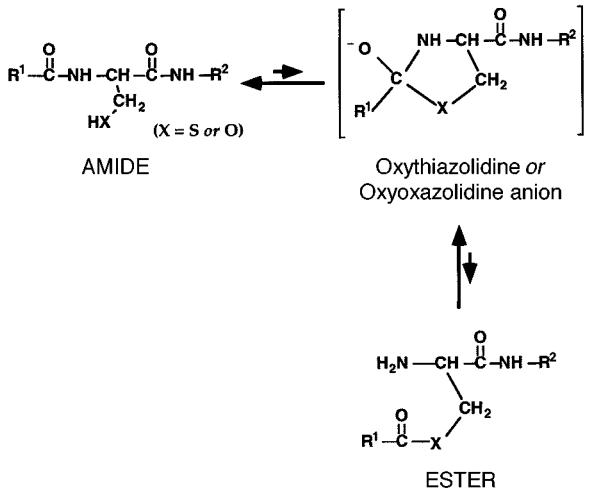


Fig. 4 N→O and N→S acyl rearrangement, with the postulated hydroxythiozolidine or hydroxyoxazolidine intermediate

Cys residues and hydroxythiazolidine intermediates can also occur, these have not been studied. Acid-induced N→S acyl shifts are less frequent in proteins than N→O acyl shifts owing to the low levels of Cys in proteins relative to Ser and Thr and the equilibrium is even more unfavorable, thioesters being about 50-fold less stable than oxygen esters.[13] Moreover, at pH 5.0, the lowest pH at which amino groups can be trapped by acylation following N→O acyl shifts,[12] the rates of S→N acyl rearrangements are extremely rapid, with half-lives of only 24 s,[8] precluding chemical trapping of the thioester product for quantitation under neutral conditions.

3.2 N→O and N→S acyl rearrangements in protein splicing

The occurrence of N→O and N→S acyl rearrangements in protein splicing was demonstrated using an *in vitro* splicing system involving the intein from the DNA polymerase of the hyperthermophilic archeon, *Pyrococcus* sp. GB-D.[7] Because thioesters react much more rapidly than oxygen esters with nitrogen nucleophiles at neutral pH, protein splicing precursors in which the Ser residue of interest was replaced by Cys were constructed and purified. In the presence of 0.25 M hydroxylamine or 0.1 M ethylenediamine at pH 6 or higher, these constructs undergo rapid cleavage at the upstream splice junction, consistent with the aminolysis of a thioester (Fig. 5). The site of hydroxylaminolysis was identified by analysis of the C-terminus of the polypeptide cleavage products. Comparison of the C-terminal peptide hydroxamate with synthetic peptide hydroxamates with respect to chromatographic mobility, colorimetric assay, amino acid composition, and high resolution mass spectrometry showed that the hydroxylamine-sensitive site in the splicing precursor is the peptide bond adjacent to the serine residue at the upstream splice junction. These studies were subsequently extended to the VMA intein of *S. cerevisiae*, in which Cys occurs naturally at the upstream splice junction, by studying the hydroxylaminolysis of intein variants that were prevented from undergoing protein splicing by the replacement of essential amino acid residues at the intein C-terminus.[11] Studies on the N→S acyl rearrangements of such mutant derivatives of the *S. cerevisiae* VMA intein showed that the thioester intermediate can also be cleaved by transesterification with a large excess of low molecular weight thiols such as cysteine or 1,4-dithiothreitol (Fig. 5), an observation that made possible the development of a protein expression/purification system with self-cleavable affinity tags.[14]

The fact that the thioester product of N→S acyl shifts cannot be trapped chemically at neutral pH when generated by acid-induced rearrangements of proteins (see Section 3.1) but can be trapped by hydroxylamine or thiols when generated catalytically during protein splicing suggests that the position of the N↔S equilibrium differs in the two cases. This thermodynamic difference obviously cannot be the result of catalysis. Rather, its existence implies that the bond that undergoes the N→S or N→O acyl shift in protein splicing is not a typical peptide bond but is under strain that can be relieved by the rearrangement. Support for this idea was provided by the crystallographic analysis of the GyrA intein from *Mycobacterium xenopi*, which showed the peptide bond in question to be in the uncommon *cis* configuration.[15] The energy of *cis*-peptide bonds is about 5 kcal mol^{-1} higher than that of *trans*-peptide bonds, making the equilibrium constant of N→S acyl rearrangements in which a *cis*-peptide bond is rearranged to an ester bond about 4000 times larger. Furthermore, a *cis*-peptide precursor resembles the cyclic hydroxythiazolidine intermediate more closely than a *trans*-peptide precursor with respect to atomic coordinates (Fig. 6), suggesting that N→S acyl shifts involving precursors with a *cis*-peptide bond have lower activation energies and thus may have a kinetic as well as a thermodynamic advantage over the corresponding reactions in an all-*trans* peptide. It will be of great interest to learn whether the peptide bond linking the intein to the N-extein is always in the *cis*-configuration.

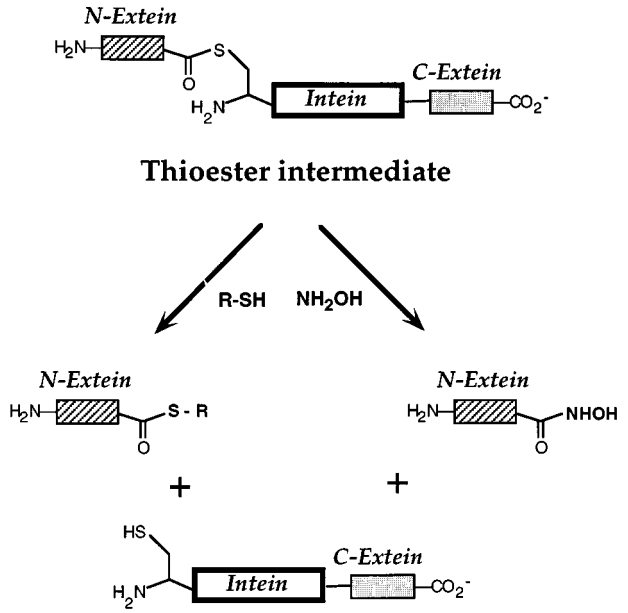


Fig. 5 Trapping of thioester intermediates by nucleophilic displacement by hydroxylamine or an excess of thiol

Studies on the role of the C-terminal amino acid of the N-extein in protein splicing involving the *S. cerevisiae* VMA intein support the notion that the geometry of the peptide bond undergoing the N→S acyl shift is important.[16] When Val, Leu, Glu, Phe, Tyr, Trp, Lys, or Thr participate in this peptide bond, substantial proportions of thioester intermediate can be trapped by denaturation of the unspliced precursor with 8 M urea, indicating that the N↔S equilibrium is far displaced towards the thioester, with equilibrium constants at pH 7.6 as high as 10 (for the Leu-Cys bond). The ester intermediate was identified by thiol-induced cleavage using a large excess of dithiothreitol and is not observed when the scissile bond involves Gly, Ala, Ile, Ser, Gln, His, Cys, Asn, and Pro. The observation that the N↔S equilibrium for a splicing junction comprising Leu-Cys favors thioester formation, whereas an Ile-Cys junction exists mainly in the amide precursor form, serves to illustrate the subtle structural factors that drive protein splicing.

Little is known about the catalytic groups involved in the N→S acyl rearrangement beyond what can be inferred from the crystal structure of the GyrA intein from *M. xenopi*.[15] This structure was obtained using an intein analog in which Cys-1 was replaced by Ser and which carried a single Ala residue as

the N-extein but lacked a C-extein. (To avoid confusion, we will refer to Cys-1 in the discussion that follows, although the crystal structure was obtained with the Ser-1 analog of the GyrA intein.) A key element in the reaction appears to be the conserved intein Block B, which contains the sequence Thr-x-x-His, located about 75 residues from the upstream splice junction (see Fig. 2).[2,17] A similar Thr-x-x-His motif is found in the hedgehog proteins (see Section 3.3.1). In the GyrA intein from *M. xenopi*, this conserved sequence is Thr-Ala-Asn-His and comprises intein residues 72–74. An H-bond between Thr-72 and the amide N of Cys-1 as well as steric constraints force the scissile peptide bond into the *cis* conformation (Fig. 6). The thiol group of Cys-1 is poised for nucleophilic attack on the carbonyl of the scissile peptide bond. (The nucleophilic attack probably involves the thiolate anion of Cys-1, which is replaced by a Ser hydroxy in the crystal structure.) The resulting tetrahedral intermediate, an oxythiazolidine anion, is stabilized by the side chain hydroxy of Thr-72 and the side chain amide of Asn-74, analogous to the stabilization of the tetrahedral intermediate in serine protease catalysis by the so-called oxyanion hole. The imidazolium group of His-75 is capable of donating a proton to the amino N of the scissile bond to promote the collapse of the oxythiazolidine anion to the thioester product (Fig. 6).

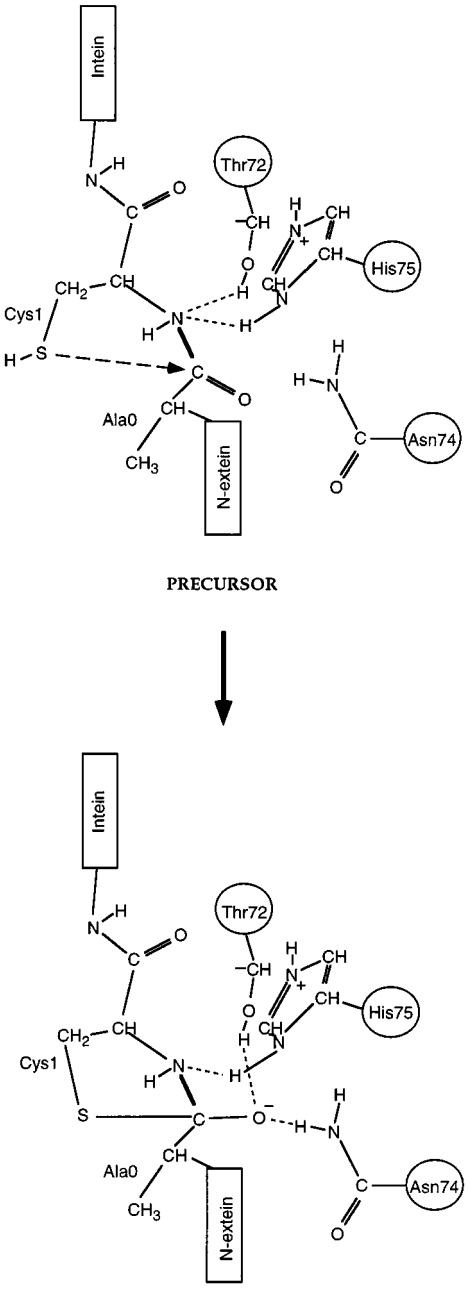


Fig. 6 Mechanism of N→S acyl rearrangement inferred from the crystal structure of an analog of the GyrA intein from *M. xenopi*. The diagram shows the amino acid residues at the N-terminal splice junction, with Cys replacing Ser at the intein N-terminus of the crystal structure, and the side chains of the amino acids thought to be involved in catalysis. The positions of the C, N, and O atoms are based on the atomic coordinates of the precursor determined by Klabunde *et al.*[15] and were modeled by the RasMol program. The positions of H atoms and the atoms of the oxythiazolidine anion are estimates.

3.3 Other self-catalyzed N→O and N→S acyl rearrangements

Nucleophilic attack by a Ser or Cys side chain on the adjacent peptide bond, leading to N→S or N→O acyl shifts, is not only seen in protein splicing but also in many other self-catalyzed protein rearrangements (Fig. 7). The broad distribution and diverse biological functions of the proteins involved suggest that such self- catalyzed rearrangements arose relatively early in evolution.

3.3.1 Autoprocessing of hedgehog proteins

The hedgehog proteins are signaling proteins that function in embryonic patterning of multicellular animals, from nematodes to mammals. A typical example is the hedgehog protein from *Drosophila melanogaster*. A 45 kDa protein precursor is secreted and then undergoes self-catalyzed processing involving polypeptide chain cleavage adjacent to a Cys residue to yield a 25 kDa C-terminal fragment (Hh-C or autoprocessing domain) and a 20 kDa protein (Hh-N or signaling domain) with a C terminus that is esterified with cholesterol. The esterified Hh-N signaling domain, which is lipophilic owing to its C-terminal cholesterol moiety, interacts with specific cell-surface receptors and is responsible for developmental signaling.[18] The autocleavage in hedgehog proteins occurs adjacent to a highly conserved Cys residue. This Cys is followed by a 12-amino acid sequence motif that resembles the conserved Block A motif of self-splicing proteins (see Fig. 2),[19] suggesting mechanistic similarities between protein splicing and hedgehog autoprocessing.[20] The mechanism of hedgehog protein autoprocessing was studied using a hedgehog protein analog in which a hexahistidine sequence replaced most of the Hh-N domain, thereby allowing for facile purification of the precursor protein by affinity chromatography on Ni^{2+} resin after expression in *E. coli*. Evidence for a thioester intermediate came from the observation that hydroxylamine and thiols promote cleavage of the precursor, yielding the hydroxamate or thiol ester of Hh-N (see Fig. 5).[20,21] Replacement of His-329, which resides in a conserved domain homologous to the Block B motif of self-splicing proteins (see Fig. 2), blocks the autoprocessing reaction, suggesting that Block B plays a role in the acyl rearrangements leading to the ester intermediates in protein splicing and hedgehog protein autoprocessing. Comparison of the amino acid sequences of hedgehog autoprocessing domains (Hh-C) and protein splicing elements shows considerable homology,[22–24] which is preserved on the level of protein structure,[24] suggesting a close evolutionary relationship between protein splicing elements and hedgehog autoprocessing domains.

3.3.2 Autocleavage of glycosylasparaginase precursors

Glycosylasparaginase hydrolyzes N4-(β-*N*-acetylglucosaminyl)-l-aspartic acid and related glycans, and its deficiency in humans leads to aspartylglycosaminuria, a genetic disorder of glycoprotein degradation. The human and *Flavobacterium meningosepticum* glycosylasparaginases have a high degree of amino acid sequence and structural homology, both being composed of α and β subunits, which are encoded by a single

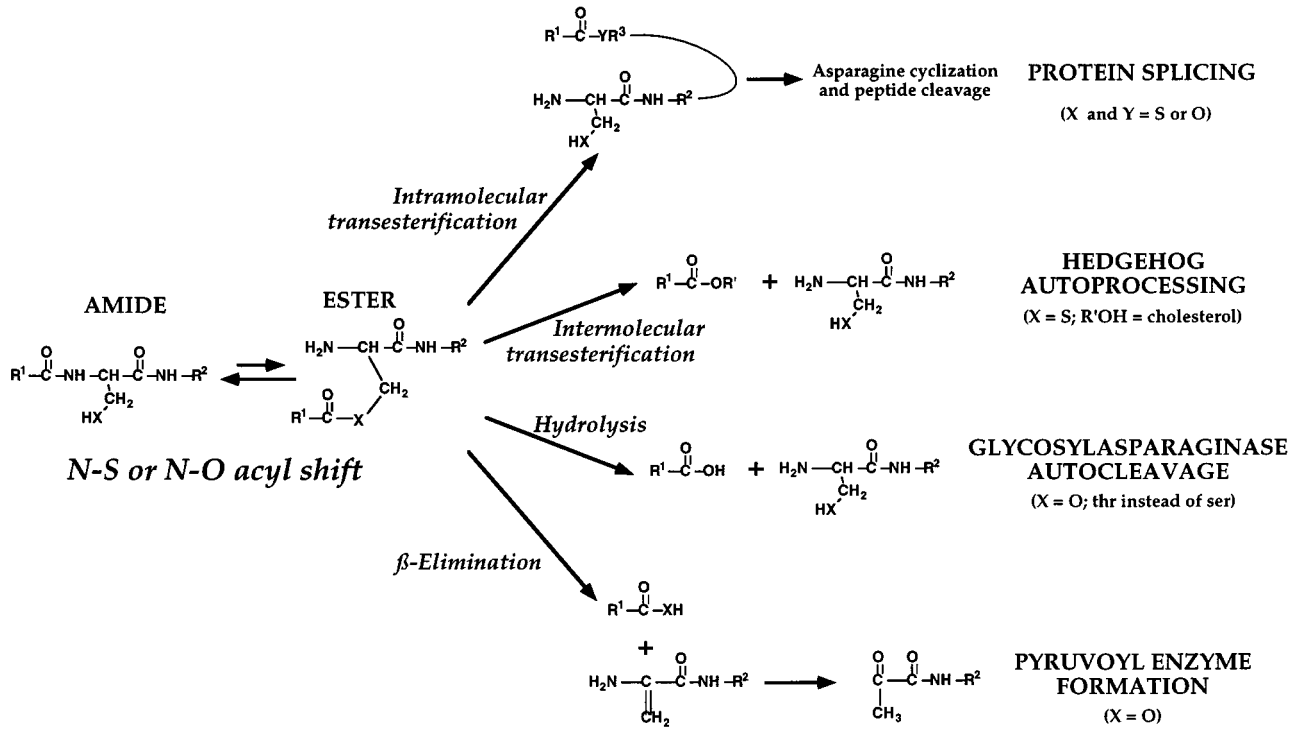


Fig. 7 The role of N→O and N→S acyl rearrangements in self-catalyzed protein rearrangements

gene. The primary gene product, an inactive polypeptide precursor, is converted to the active enzyme by self-catalyzed cleavage adjacent to a Thr residue to yield the α and β subunits, the latter with an N-terminal Thr.[25] Replacement of this Thr residue in bacterial glycosylasparaginase with amino acids other than Ser or Cys completely prevents autocleavage, whereas replacement with Cys or Ser greatly reduces its rate.[26] The use of affinity-tagged recombinant proteins in which the critical Thr residue is replaced by Cys, in conjunction with the discovery that glycine severely inhibits autocleavage, made possible the purification of glycosylasparaginase precursors by affinity chromatography and the study of the autocleavage mechanism. Evidence for the involvement of an ester intermediate in autocleavage comes from the observations that the cleavage of the Thr→Cys mutant protein is promoted by the nucleophile hydroxylamine and inhibited by the thiol-blocking reagent iodoacetamide.[26,27] These studies also provided important insights into the mechanism of the N→O acyl shift leading to ester formation, in which the tripeptide His-Asp-Thr at the potential cleavage site plays a critical role. The imidazole side chain of His-150 acts as proton acceptor to facilitate the nucleophilic attack of the Thr-152 hydroxy group on the peptide carbonyl of the Asp-Thr peptide bond [Fig. 8(a)]. The Asp-151 side chain is anchored in the partially formed substrate binding pocket of the enzyme precursor in a conformation that strains the Asp-Thr peptide bond. The relief of this strain upon N→O acyl rearrangement serves to compensate for the unfavorable equilibrium constant associated with N→O acyl shifts. Binding of the Asp-151 side chain in the incipient substrate binding site is stabilized by ionic interactions with the guanidino group of Arg-180. Owing to its small size and ionic interactions of its carboxy and amino groups with Arg-180 and Asp-183, respectively, free glycine can also bind strongly to the same site. The important role of the binding of the Asp-151 β-carboxylate moiety in the incipient substrate binding pocket is supported by three lines of evidence: (i) the strong inhibition of autocleavage by free glycine, which competes with the Asp side chain for the incipient substrate binding site; (ii) the suppression of autocleavage by mutations in which the Asp-151 residue is replaced by other amino acids; and (iii) the activation of autocleavage in an Asp-151→Gly mutant by the addition of glycine, which, by substituting for the missing Asp-151 side chain in the substrate binding pocket, restores the original, strained protein conformation.[27]

An interesting aspect of the autocleavage of the glycosylasparaginase precursor is that the N-terminal Thr residue of the β subunit generated by this process is essential for enzyme activity. In this respect, glycosylasparaginase resembles a group of other amidohydrolases, termed N-terminal nucleophile amidohydrolases, which include glutamine-phosphoribosylpyrophosphate amidotransferase, penicillin acylase, and the proteasome. These enzymes have similar three-dimensional structures and their activities depend on an N-teminal Ser, Thr, or Cys residue that is generated by the cleavage of a precursor protein, probably through self-catalysis involving N→O or N→S acyl rearrangements.[28] It is interesting that the nucleophilic amino acid in the N-terminal nucleophile amidohydrolases plays quite similar roles mechanistically in the autoprocessing reaction and in the catalysis of amide hydrolysis by the mature enzyme. This is illustrated in Fig. 8 for glycosylasparaginase.[27,29] The possibility of a similar dual role for the N-terminal nucleophilic residue of inteins in the first and third steps of protein splicing is discussed in Section 5.2. In spite of these possible functional similarities, the N-terminal nucleophile amidohydrolases exhibit no structural or sequence similarity to protein splicing elements.

3.3.3 Autocleavage leading to pyruvoyl enzyme formation

The activity of pyridoxyl phosphate-independent bacterial amino acid decarboxylases and reductases depends on N-terminal pyruvoyl prosthetic groups. The pioneering work of E.E. Snell and coworkers (reviewed in ref. 30) established that the enzyme-bound pyruvate is generated through the self-catalyzed cleavage of a protein precursor. In the case of the histidine decarboxylase from *Lactobacillus* 30a, processing of the proenzyme involves cleavage of the peptide bond beween Ser-81 and Ser-82, coupled to the conversion of Ser-82 to an N-terminal pyruvate residue by β-elimination. The first suggestion that this process involves an ester intermediate came from isotopic labeling experiments, which showed that ^{18}O from the side chain of Ser-82 is transferred to the Ser-81 carboxy group in the course of the cleavage reaction. The putative ester

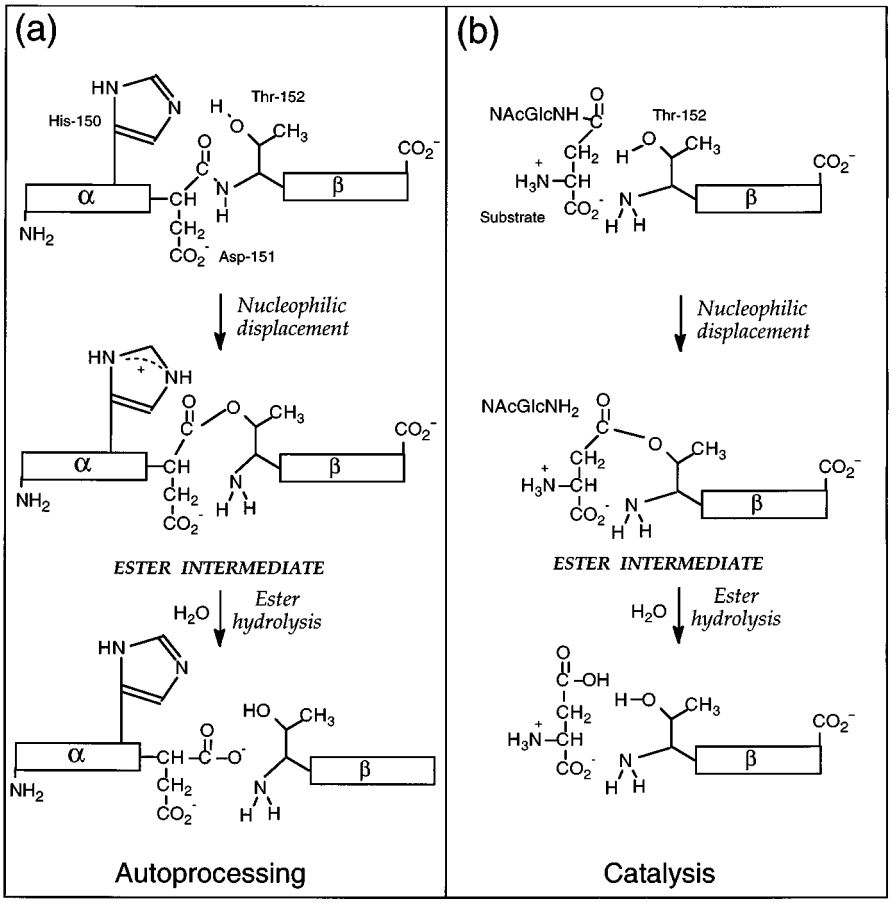


Fig. 8 Comparison of the role of Thr-152 in (a) autocleavage of the glycosylasparaginase precursor and (b) catalysis of amide hydrolysis by the mature enzyme, based on the studies of Guan and coworkers.[27,29] Only the residues at the catalytic site are shown, with the α and β domains of the protein indicated by rectangular boxes. Abbreviation: NAcGlc, β-*N*-acetylglucosaminyl.

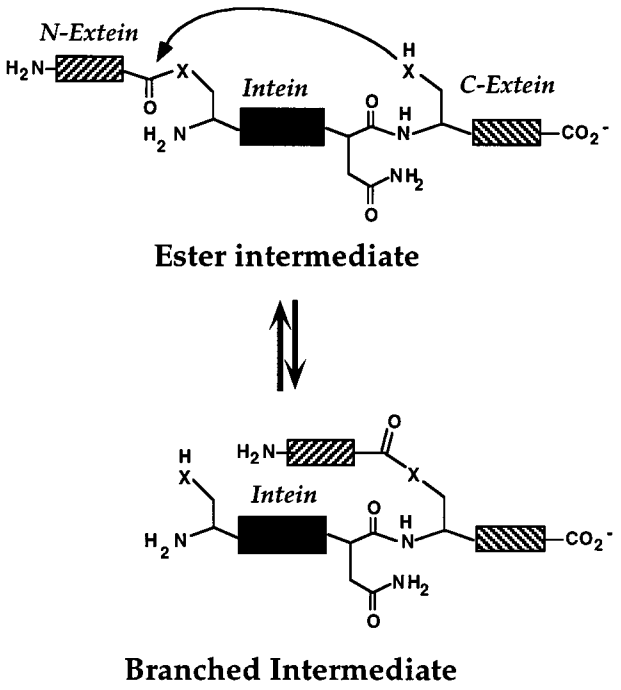
intermediate could be trapped as the hydroxamate after replacing the relevant Ser residue with Cys,[31] owing to the much greater susceptibility of thioesters than oxygen esters to attack by nitrogen nucleophiles. Direct evidence for an ester intermediate in the autoprocessing of pyruvoyl enzyme precursors came from X-ray crystallographic studies on a pyruvoyl enzyme, aspartate decarboxylase from *E. coli*.[32] Aspartate decarboxylase is synthesized as a single polypeptide chain, π, that undergoes self-cleavage to generate α and β subunits, with the β subunit carrying an N-terminal pyruvate residue. The active enzyme crystallizes as a tetramer with the structure $(\alpha\beta)_3(\pi)$, which has only 3 pyruvoyl groups, presumably owing to negative cooperativity in the processing reaction. The 2.2 Å electron density map is inconsistent with a peptide bond at the potential cleavage site between Gly-24 and Ser-25 of the unprocessed π subunit but shows a satisfactory fit with a Gly-Ser ester bond, suggesting that an N→O acyl shift had occurred in the π subunit to yield a stable ester. The stability of the peptide ester in the crystals of the aspartate decarboxylase tetramer, grown at pH 7.5, is remarkable. Two factors may be responsible for the stabilization of the ester: (i) the protonation of the free serine amino group by Tyr-58, and (ii) the relief of steric strain within the 5-amino acid loop containing Gly-24 and Ser-25, which is unusually short for a loop linking two β strands.[32]

The aspartate decarboxylase from *E. coli* shows neither sequence nor structural homology with other pyruvoyl enzymes, nor with glycosylasparaginases, hedgehog proteins, or protein splicing elements. This suggests that, except for protein splicing elements and hedgehog proteins, which show clear signs of an evolutionary relationship, self-processing proteins that take advantage of N→O or N→S acyl shifts to effect peptide bond cleavage have evolved independently. An independent origin of these self-processing events is also suggested by the diverse mechanisms for overcoming the unfavorable equilibrium of N→O and N→S acyl rearrangements. Although based on the same underlying principle— destabilization of the peptide bond that is to undergo the acyl rearrangement—the various mechanisms by which this is achieved include a *cis*-peptide bond (GyrA intein), anchoring of an amino acid side chain in an incipient substrate binding pocket (glycosylasparaginase), and tertiary structure constraints (aspartate decarboxylase).

4 Transesterification (Scheme 2)

Intramolecular transesterifications are relatively rapid and facile reactions, involving both effective nucleophiles and favorable equilibria. In protein splicing, transesterification occurs either between a thiol and a thioester, an alcohol and an oxygen ester, or an alcohol and a thioester, but never between a thiol and an oxygen ester.[2] In the first two cases, the equilibrium constant at neutral pH is near unity; in the alcohol–thioester reaction it is about 50.[13] Accordingly, in 40% of the known inteins where an N→S acyl shift is followed by transesterification with a Ser or Thr side chain at the downstream splice junction, the transesterification step will partially compensate for the unfavorable equilibrium of the N→S acyl shift. However, even when the equilibrium constant for transesterification is close to one, as in the alcohol–oxygen ester or thiol–thioester reactions, kinetic stabilization occurs as discussed below.

Evidence for the occurrence of transesterification in protein splicing came from the characterization of a transient interme-



Scheme 2

diate that occurs in the *in vitro* splicing reaction involving the DNA polymerase intein from *Pyrococcus* sp. GB-D discussed earlier.[3] The intermediate, which appears at 37 °C with a half-time of about 15 min and migrates abnormally slowly upon polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate, was found to have two N-termini, corresponding to those of the N-extein and the intein. The unusual branched structure of this intermediate accounts for its abnormal electrophoretic mobility and implies that two polypeptides with free N-termini are linked to a single Ser residue, one to its α-amino group and the other to its β-hydroxy group. In order to determine which of the two N-terminal polypeptides, the N-extein or the intein, is esterified with the Ser side chain hydroxy group, protein splicing was interrupted by adding 6 M guanidinium chloride when an appreciable amount of the branched intermediate had accumulated and the mixture was incubated at pH 9.0 and 65 °C, at which amide bonds are stable but esters are hydrolyzed. Under these conditions, free N-extein is released from the branched intermediate, indicating that it had been linked by an ester bond, consistent with the reaction pathway outlined in Fig. 3.[9]

The intein from the DNA polymerase of *Pyrococcus* sp. GB-D has Ser residues at both splice junctions and the transesterification reaction should therefore have an equilibrium constant of one and be freely reversible. Since the transesterification equilibrium is coupled to the pH-dependent N→O acyl rearrangement equilibrium, the reversibility of transesterification can be experimentally verified by controlling the pH. Raising the pH to 10 causes the branched intermediate to revert to the linear precursor form with a $t_{1/2}$ of about 10 min; readjusting the pH to 6 leads to regeneration of the branched intermediate and subsequent protein splicing, indicating full reversibility.[3] It is interesting to compare the $t_{1/2}$ for the reversal of transesterification with the $t_{1/2}$ for the O→N acyl shift, which is 27 min at pH 6.0 and declines precipitously with increasing pH (to about 2 min at pH 7.0 and an unmeasurably small value at higher pH).[8] The rate at which the branched intermediate reverts to the linear precursor is several orders of magnitude slower than the rate of the O→N shift *per se*,[3] suggesting that the transesterification reaction provides significant kinetic stabilization of ester intermediates. Such kinetic stabilization may play an important role when protein splicing involves an N→S acyl rearrangement, because the estimated $t_{1/2}$ for the S→N acyl shift at pH 7 is only about 0.1 s.

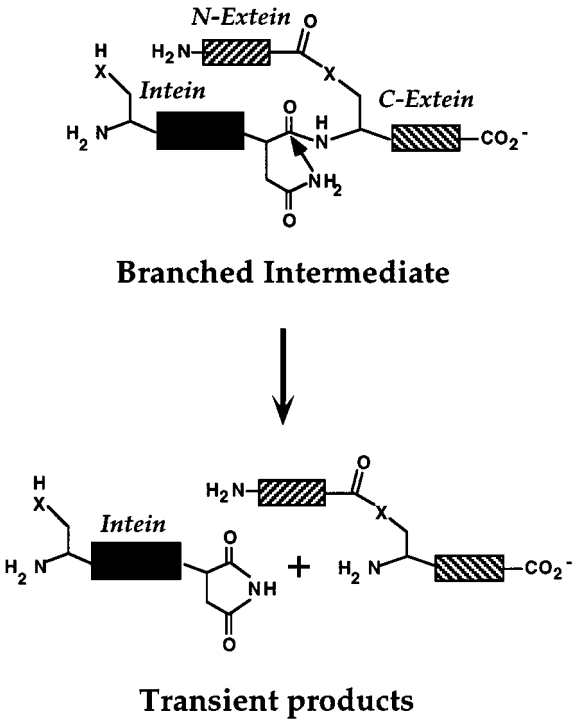
The crystal structures of the *S. cerevisiae* VMA intein[33] and the *M. xenopi* GyrA intein[15] were determined with modified proteins lacking the C-terminal Cys or Thr residues that are involved in transesterification, and little can therefore be said about the amino acid side chains that might assist the transesterification reaction. However, even when the position of the C-terminal Thr is approximated by modeling, no amino acid side chain that might provide catalytic assistance can be identified.[15] It is possible that the thiol of the Cys that participates in transesterification in the yeast VMA intein and many other inteins has an unusually low pK_a and is therefore an effective nucleophile between pH 6 and 7, where splicing is optimal. The observation that a disulfide bond between the Cys residues at the splice junctions of the *M. tuberculosis* RecA intein forms rapidly under very mildly oxidizing conditions during the reconstitution of a functional protein splicing element from N- and C-terminal fragments[34,35] supports this notion and also indicates that the spatial arrangement of the Cys thiols at the two splice junctions is well suited for transesterification.

Although esters and thioesters play an important role in protein metabolism, ranging from protein degradation to the defensive functions of the complement system and α-macroglobulins and the biosynthesis of lipids and peptide antibiotics, there are no documented examples of transesterification reactions that occur within a single polypeptide chain other than the reactions that participate in protein splicing. It seems, therefore, that the self-catalyzed intramolecular transesterifications observed in protein splicing represent a unique evolutionary adaptation for rearranging polypeptide backbones efficiently and without the necessity of an external energy source. Hedgehog proteins, which bear a close evolutionary relationship to protein splicing elements, may mediate an intramolecular transesterification similar to that seen in the second step of protein splicing.[23] These proteins undergo an N→S acyl shift, followed by intermolecular transesterification with the 3-OH group of cholesterol catalyzed by the autoprocessing domain.[18] It has been suggested that the reaction with cholesterol occurs after prior intramolecular transesterification with a Cys residue found in a highly conserved region near the C-terminus of the autoprocessing domain, thereby producing a branched ester intermediate similar to that seen in protein splicing.[23]

5 Peptide cleavage coupled to asparagine cyclization (Scheme 3)

5.1 Spontaneous cyclization of asparagine in peptides and proteins

The cyclization of Asn residues is frequently observed in peptides and proteins (reviewed in ref. 36). This reaction can follow either of two routes: (i) attack by the peptide bond N on the carbonyl C of the Asn β-amide, leading to deamidation [Fig. 9(a)], or (ii) attack by the Asn β-amide N on the carbonyl C of the peptide bond, leading to peptide bond cleavage [Fig. 9(b)]. Asn cyclization can occur at extremes of pH and can become a significant side reaction during the deprotection of synthetic peptides under acidic conditions, but it is also observed at neutral pH, albeit at much lower rates. A critical determinant of the propensity for Asn cyclization is the conformation of the peptide bond that participates in the reaction, particularly the dihedral angle ψ, as well as the orientation of the Asn side chain, defined by the dihedral angle χ.[36] The optimal ψ (+120° or −120°) and χ angles (+90° or −90°) for Asn cyclization are not those favored in typical polypeptides. As a result, not all Asn residues in proteins are equally susceptible to cyclization. Asn residues adjacent to Gly, which allows a much greater range of dihedral angles than amino acids with bulky side chains, are particularly prone to



Scheme 3

cyclization. The most common mode of Asn cyclization in proteins is the one leading to deamidation. Even though its rate is slow under physiological conditions, deamidation can be a significant reaction in long-lived proteins and can affect protein function owing to racemization and ring opening, which leads to the generation of Asp and iso-Asp residues (Fig. 9).[37] Asn cyclization leading to peptide bond cleavage occurs at high temperatures in peptides in which Asn is followed by an amino acid with a bulky side chain,[37] suggesting that steric constraints for attack by the side chain amide N are different than for attack by the peptide bond N. This mode of peptide cleavage has also been found in α-crystallin, an unusually long-lived protein.[38]

5.2 Asparagine cyclization in protein splicing

The presence of an Asn residue at the C-terminus of all inteins provides the opportunity for the excision of the intein by Asn cyclization. That such a reaction indeed occurs was demonstrated by analysis of the C-terminal residue of the excised DNA polymerase intein from *Pyrococcus* sp. GB-D. The excised intein was cleaved with CNBr at a Met residue that had been introduced near the C-terminus of the intein by a conservative amino acid replacement. The resulting C-terminal peptide was purified by reverse phase high performance liquid chromatography, analyzed both by mass spectrometry and a specific color test for succinimides, and compared with the corresponding chemically synthesized peptide.[6,9] Similar experiments confirmed that the excised VMA intein of yeast also has a C-terminal aminosuccinimide residue, suggesting that the same cleavage mechanism functions in the intein of hyperthermophiles and mesophiles, regardless of whether protein splicing involves oxygen or thioesters.[11]

The mechanism of Asn cyclization and the attendant cleavage of the peptide bond linking the intein to the C-extein is not yet clear. Evidence that the adjacent His residue plays an important role comes from the observation that its replacement with other amino acids leads to the accumulation of the branched intermediate.[10] Examination of the crystal structure of the *M. xenopi* GyrA intein, extended by modeling a C-terminal Thr, shows that the imidazole moiety of His-197 can donate a proton

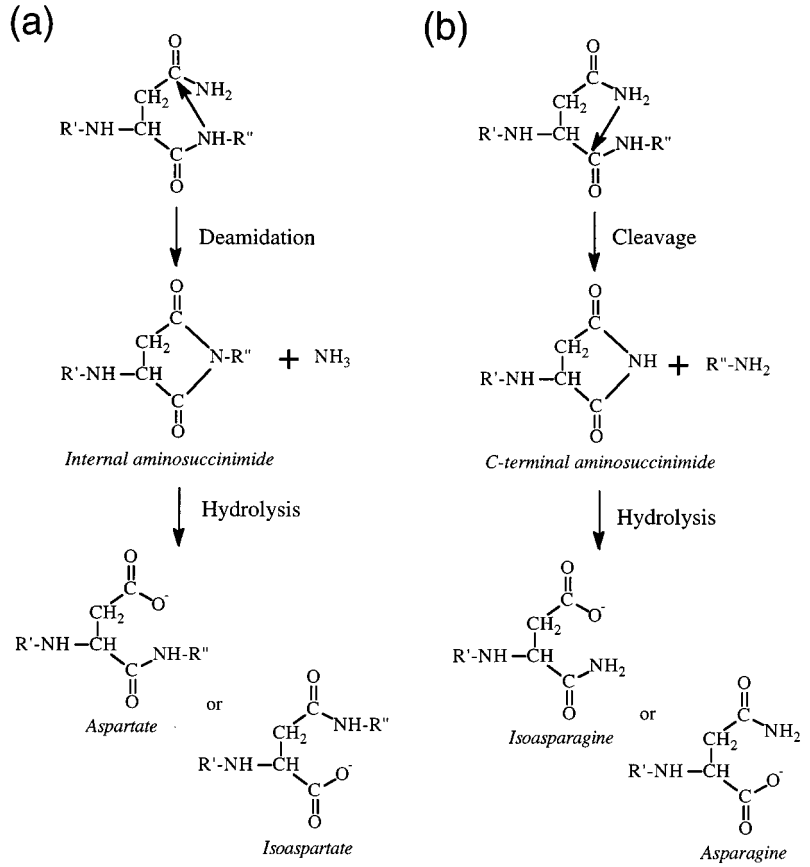


Fig. 9 Alternate modes of cyclization of Asn residues in peptides and proteins, leading either to (a) deamidation or (b) peptide bond cleavage. (Reproduced from ref. 6. Copyright 1995, American Chemical Society).

to the amido N of the scissile peptide bond to facilitate bond cleavage. Although His-197 can also form a H-bond with the carbonyl O of the Asn side chain, no clear mechanism for enhancing the nucleophilicity of the Asn amide N is apparent. An alternative mechanism for Asn cyclization involving isomerization to an imide and the attack of the side chain carbonyl O on the peptide carbonyl C has been suggested.[36] This mechanism would yield an isoimide rather than a succinimide but offers no advantages with respect to the catalytic groups available at the protein splicing active center. On the other hand, an intriguing possibility that avoids the problem posed by the poor nucleophilicity of the Asn side chain involves the participation of the nucleophilic side chain of the N-terminal Cys or Ser residue of the intein, which would function like the N-terminal nucleophile in amide hydrolysis by the mature N-terminal nucleophile amidohydrolases [see Fig. 8(b)].[28]

This alternative mechanism for Asn cyclization, which was suggested by Dr A.L. Nussbaum, Harvard Medical School (personal communication), is illustrated in Fig. 10. The Cys or

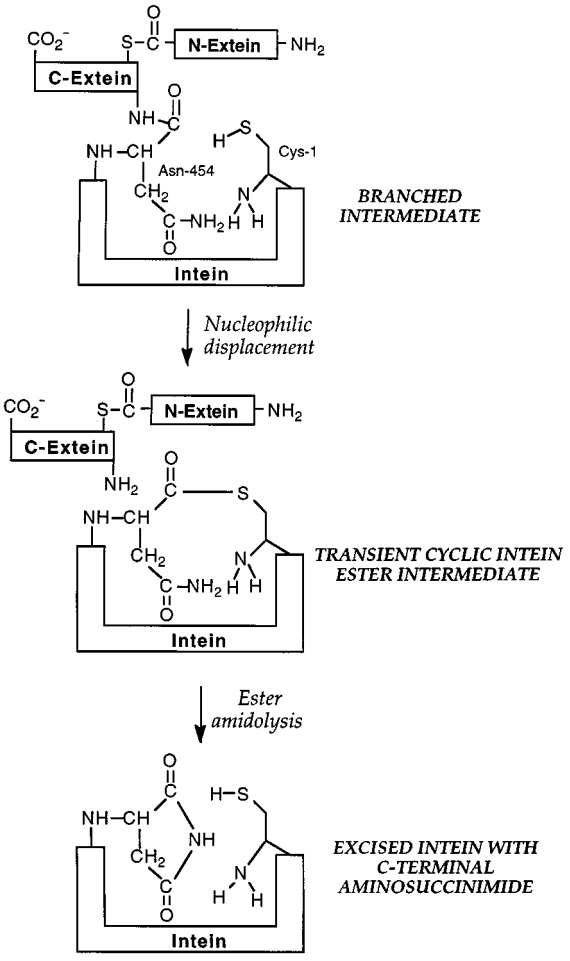
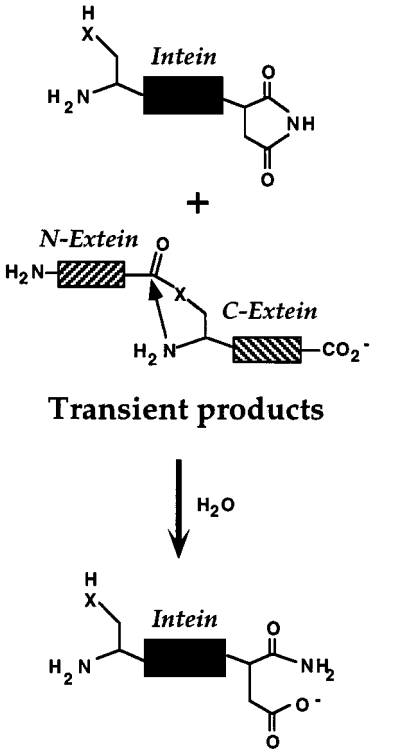


Fig. 10 Postulated mechanism for Asn cyclization in Step 3 of protein splicing with participation of the nucleophilic residue at the N-terminus of the intein and an Asn ester intermediate

Ser at the N terminus of the intein, released as an N-terminal nucleophile by the N→S or N→O acyl shift and transesterification reactions, attacks the peptide bond linking the C-terminal Asn to the C-extein in a manner analogous to catalysis by N-terminal nucleophile amidohydrolases. This leads to the liberation of the C-extein and the formation of a macrocylic intermediate in which the N- and C-termini of the intein are linked by an ester bond. The esterified C-terminal Asn residue in the macrocyclic intermediate then cyclizes by attack of the amide N on the ester carbonyl C, leading to ester hydrolysis and a linear excised intein with a C-terminal aminosuccinimide residue. Experimental support for this mechanism comes from studies on the effect of amino acid substitutions on the splicing of chimeric proteins in which an intein has been inserted in a foreign context. When plasmids encoding such chimeric proteins are expressed in *E. coli*, one observes not only protein splicing but also hydrolytic side reactions at either splice junction, especially when normal protein splicing is attenuated by amino acid substitutions.[3,9–11,39] These cleavage reactions have been shown to involve the same basic mechanisms as protein splicing[11] and thus offer the opportunity to study the effect of amino acid substitutions on individual steps of the protein splicing mechanism.[10,16] Using such experimental systems, it was found that the replacement of the N-terminal nucleophilic residue of the intein with Ala causes at least a 25-fold reduction in the rate of Asn cyclization and attendant cleavage of the C-terminal splicing junction.[10] Moreover, when the transesterification step in precursors involving the yeast VMA intein is prevented by replacing the Cys residue at the downstream splice junction with Ala and the cyclization of the C-terminal Asn-454 is attenuated by replacing the adjacent His-453 with Gln, Asn cyclization and peptide bond cleavage at the downstream splice junction requires prior thiol-induced cleavage at the upstream splice junction.[16] These results show that Asn cyclization coupled to peptide bond cleavage is much enhanced by a free intein N-terminus consisting of a nucleophilic amino acid residue. With the caveat that these observations were made in experimental systems where protein splicing was compromised by amino acid substitutions and may not be representative of normal inteins, a plausible interpretation is that Asn cyclization is preceded by a nucleophilic displacement at the peptide bond carbonyl of the downstream splice junction involving an N-terminal nucleophile, analogous to amide hydrolysis by N-terminal nucleophile amidohydrolases [compare Figs. 8(b) and 10]. This would lead to a macrocyclic peptide ester involving the Asn carboxy group, which would be resolved into a linear polypeptide by Asn cyclization (Fig. 10). Owing to the free rotation of ester bonds, cyclization of Asn in the postulated ester intermediate would be subject to much less steric constraint than cyclization of the same Asn linked by a planar peptide bond to an amino acid whose side chain is esterified with the N-extein. On the other hand, cleavage at the C-terminal splice junction proceeds slowly in the absence of N-terminal cleavage[3,10,39] or when the N-terminal nucleophile is replaced by Ala,[10] suggesting that the direct cyclization of Asn can also occur at a moderate rate. Nevertheless, the notion that the major route to Asn cylization involves attack by an N-terminal nucleophile to form a macrocylic ester intermediate is attractive because it would assure that C-terminal cleavage can occur only after cleavage at the N-terminal splice junction and the formation of the branched intermediate. Such a mechanism would assure that the steps in protein splicing occur sequentially and thereby minimize side reactions leading to abortive cleavage.

6 Uncatalyzed finishing reactions (Scheme 4)

With the cyclization of the C-terminal Asn residue and the attendant cleavage of the peptide bond between the intein and the C-extein, protein splicing has been achieved in the sense that the intein has been excised and the exteins have been linked together. However, the excised intein contains an unnatural C-terminal residue, aminosuccinimide, and the exteins are linked by an unnatural ester bond. The completion of the splicing process requires the elimination of these unnatural features through some finishing reactions. Owing to the fact that the intein, which catalyzed the earlier steps in protein splicing, has now been excised and lacks specific affinity for the exteins, the finishing reactions, by necessity, have to be spontaneous reactions that can proceed rapidly in the absence of catalysts.



Scheme 4

6.1 Hydrolysis of C-terminal aminosuccinimides

The rate of hydrolysis of C-terminal aminosuccinimides to Asn or iso-Asn [see Fig. 9(a)] was measured using synthetic tetrapeptides corresponding to the C-terminus of the DNA polymerase intein from *Pyrococcus* species GB-D.[6] The rate of succinimide hydrolysis at 37 °C is strikingly pH-dependent, with a half-life of 350 h at pH 5.5, which declines to 17 h at pH 7.4. These rates are considerably slower than the rates of hydrolysis of N-substituted cyclic imides produced at internal positions of polypeptide chains by nucleophilic attack of the peptide bond N on the carbonyl C of the Asn β-amide [Fig. 9(a)] (summarized in ref. 6). In model peptides, the hydrolysis of C-terminal aminosuccinimide proceeds with an activation energy ($\Delta H^{\ddagger}$) of 23.2 kcal mol^{-1} and with a relatively low entropy of activation ($-T\Delta S^{\ddagger}$ = 2.1 kcal mol^{-1}).[6] The relatively high stability of C-terminal aminosuccinimides implies that, at least in mesophilic organisms, a substantial fraction of the excised inteins carry C-terminal aminosuccinimide residues. This prediction is supported by the observation that about 50% of excised inteins isolated from *E. coli* transformed with an appropriate recombinant plasmid are terminated with aminosuccinimide.[9]

6.2 S→N and O→N acyl rearrangements in peptide esters

In order to estimate the rates of S→N and O→N acyl shifts, the depsipeptides N-(9-fluorenyl)methoxycarbonyl (Fmoc)-Glu-o-Ser-Asp-Gly-Tyr-NH_2 and Fmoc-Glu-s-Cys-Asp-Gly-Tyr-NH_2 were synthesized and deprotected under acid conditions where the ester form is stable. (The designation Glu-o-Ser and Glu-s-Cys indicates that these amino acids are linked by a ester or thioester bond involving the side chains of Ser or Cys, respectively, rather than by a peptide bond.) The conversion of the deprotected depsipeptides to true peptides was estimated by reverse phase chromatographic separation of peptide esters and their amide rearrangement products in the presence of 0.1% trifluoroacetic acid, which prevents further S→N and O→N acyl rearrangement.[8] Both the rates and equilibrium constants for the S→N and O→N acyl rearrangements were found to increase rapidly with pH. At pH 6.0, the half-life of the oxygen ester, Fmoc-Glu-o-Ser-Asp-Gly-Tyr-NH_2, is 27 min; at pH 5.0, the half-life of the corresponding thioester, Fmoc-Glu-s-Cys-Asp-Gly-Tyr-NH_2, is only 24 s; at pH 7, the rates of acyl rearrangement of both esters are too fast for measurement by these techniques. The difference in acyl rearrangement rates for the oxygen esters and thioesters under identical conditions is about 1000. The activation energy ($\Delta H^{\ddagger}$) for the O→N acyl shift at pH 6.0 is only 5.2 kcal mol^{-1}; that for the S→N acyl shift at pH 4.0 is even less (4.2 kcal mol^{-1}); and the entropies of activation ($-T\Delta S^{\ddagger}$) are also very low (−700 and 1200 kcal mol^{-1}, respectively). The extremely rapid uncatalyzed rates and the virtual irreversibility of S→N and O→N acyl rearrangements at neutral pH makes these acyl shifts ideal finishing reactions that rapidly drive protein splicing to completion under physiological conditions, even without catalysts.

7 Conclusions

Protein splicing is a remarkable process in which nonfunctional polypeptides undergo complex and highly specific rearrangements to yield functional proteins. It may therefore seem surprising that a review of the chemical reactions that underlie protein splicing reveals only reaction types that are already well known to the protein chemist. However, what is remarkable about protein splicing is not the novelty of the chemistry but the way in which protein splicing elements organize these chemical reactions into an effective pathway for remodeling proteins in a highly specific manner with a minimum of side reactions. Indeed, it is amazing how protein splicing elements can catalyze this complex set of chemical transformations using less than 150 amino acid residues[35,40] and no external cofactors. This contrasts with other types of post-translational modification, which often require a complex enzymatic machinery: proteases, kinases, phosphatases, nucleotidyl transferases, *etc*.

Now that the chemical basis of protein splicing is well understood, the major remaining challenge is to elucidate in detail how protein splicing elements can catalyze the reactions of the protein splicing pathway with such sparse means. As summarized in this review, significant progress has already been made in this direction, but many unanswered questions remain. Even though the crystal structures of two inteins have been solved,[15,33] the proteins that were crystallized lack one or both exteins and therefore fail to reveal the most critical interactions needed for the catalytic process. Much information has been gained from studying the effect of amino acid substitutions on intein function,[10,11] but observations based on systems perturbed by mutations must be interpreted with caution. As is evident from this review, some mechanistic insights can be gained by comparing protein splicing with other self- catalyzed protein rearrangements, such as hedgehog protein autoprocessing and glycosylasparaginase autocleavage, but these serve primarily as starting points for further study. A promising experimental system for addressing structure–function relationships in protein splicing is the recently developed *in vitro trans*-splicing system based on highly truncated intein fragments.[34,35] As shown in Fig. 11, N- and C-terminal segments of the RecA intein from *Mycobacterium tuberculosis*, each about 100 amino acids long and fused to appropriate exteins, can be reconstituted into a functional protein splicing element by renaturation from 6 M urea or guanidinium

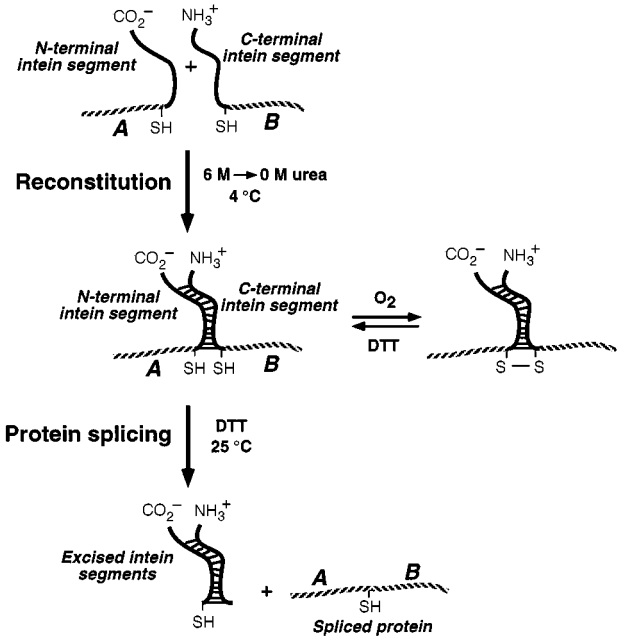


Fig. 11 Diagram illustrating the reconstitution and splicing of fusion proteins containing N- and C-terminal segments of the RecA intein, linked to exteins A and B, respectively

chloride.[34] Under mildly oxidizing conditions, the intein segments are almost quantitatively reconstituted into a non-reactive, covalent dimer, in which the side chains of the Cys residues at the splice junctions are linked by a disulfide bond. Upon treatment with disulfide reducing agents, the essential thiol groups are regenerated and protein splicing ensues. A semisynthetic protein splicing element can be generated by replacing the C-terminal intein segment with synthetic polypeptides corresponding to 35–50 of the C-terminal amino acids of the RecA intein.[35] An especially valuable feature of this experimental system is that the reconstitution of the intein fragments to yield a functional protein splicing element can be studied independently of the protein splicing process *per se*, thus allowing separate investigation of structural and catalytic aspects of the protein splicing element. Moreover, the fact that a semisynthetic protein splicing element is now available in which the C-terminal segment of the intein consists entirely of a synthetic polypeptide[35] offers the opportunity for probing the structure and function of the protein splicing active center by substituting other amino acids, including unnatural amino acids and structural probes, at specific positions of the polypeptide. Experimental approaches of this type should rapidly advance our understanding of protein splicing so that it may soon be possible to write a definitive review on the enzymatic basis of protein splicing.

8 Acknowledgements

I am deeply grateful to my collaborators at the Boston Biomedical Research Institute and New England Biolabs, who are named on our joint publications cited in this review, for their important contributions to the elucidation of the mechanism of protein splicing and for the pleasure of doing research together. I also thank Alex Nussbaum for stimulating discussions on the mechanism of asparagine cyclization, to Ken Mills and Peter Coleman for their constructive comments on the manuscript, to Thomas Klabunde for providing the atomic coordinates of the *M. xenopi* GyrA intein crystal structure prior to public release, and to Don Comb for introducing me to protein splicing. The research in my laboratory was funded by fellowships from New England Biolabs and a research grant (R01 GM55875) from the National Institutes of Health. Their research support is greatly appreciated.

9 References

1 F. B. Perler, M.-Q. Xu and H. Paulus, *Curr. Opin. Chem. Biol.*, 1997, **1**, 292.
2 F. B. Perler, G. J. Olsen and E. Adam, *Nucl. Acids Res.*, 1997, **25**, 1087.
3 M.-Q. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra and F. B. Perler, *Cell*, 1993, **75**, 1371.
4 M. Kawasaki, S.-i. Makino, H. Matsuzawa, Y. Satow, Y. Ohya and Y. Anraku, *Biochem. Biophys. Res. Comm.*, 1996, **222**, 827.
5 M. Kawasaki, Y. Satow, Y. Ohya and Y. Anraku, *FEBS Lett.*, 1997, **412**, 518.
6 Y. Shao, M.-Q. Xu and H. Paulus, *Biochemistry*, 1995, **34**, 10844.
7 Y. Shao, M.-Q. Xu and H. Paulus, *Biochemistry*, 1996, **35**, 3810.
8 Y. Shao and H. Paulus, *J. Peptide Res.*, 1997, **50**, 193.
9 M.-Q. Xu, D. G. Comb, H. Paulus, C. J. Noren, Y. Shao and F. B. Perler, *EMBO J.*, 1994, **13**, 5517.
10 M.-Q. Xu and F. B. Perler, *EMBO J.*, 1996, **15**, 5146.
11 S. Chong, Y. Shao, H. Paulus, J. Benner, F. B. Perler and M. Q. Xu, *J. Biol. Chem.*, 1996, **271**, 22159.
12 K. Iwai and T. Ando, *Methods Enzymol.*, 1967, **11**, 262.
13 W. P. Jencks, S. Cordes and J. Carriulo, *J. Biol. Chem.*, 1960, **235**, 3608.
14 S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus and M. Q. Xu, *Gene*, 1997, **192**, 271.
15 T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs and J. C. Sacchettini, *Nature Struct. Biol.*, 1998, **5**, 31.
16 S. Chong, K. S. Williams, C. Wotkovicz and M.-Q. Xu, *J. Biol. Chem.*, 1998, **273**, 10567.
17 S. Pietrokovski, *Protein Sci.*, 1994, **3**, 2340.
18 J. A. Porter, K. E. Young and P. A. Beachy, *Science*, 1996, **274**, 255.
19 E. V. Koonin, *Trends Biochem. Sci.*, 1995, **20**, 141.
20 J. A. Porter, D. P. Kessler, S. C. Ekker, K. E. Young, J. J. Lee, K. Moses and P. A. Beachy, *Nature*, 1995, **374**, 363.
21 J. A. Porter, S. C. Ekker, W. J. Park, D. P. v. Kessler, K. E. Young, C. H. Chen, Y. Ma, A. S. Woods, R. J. Cotter, E. V. Koonin and P. A. Beachy, *Cell*, 1996, **86**, 21.
22 S. Pietrokovski, *Protein Sci.*, 1998, **7**, 64.
23 J. Z. Dalgaard, M. J. Moser, R. Hughey and I. S. Mian, *J. Comput. Biol.*, 1997, **4**, 193.
24 T. M. T. Hall, J. A. Porter, K. E. Young, E. V. Koonin, P. E. Beachy and D. J. Leahy, *Cell*, 1997, **91**, 85.
25 R. Tikkanen, A. Riikonen, C. Oinonen, J. Rouvinen and L. Peltonen, *EMBO J.*, 1996, **15**, 2954.
26 C. Guan, T. Cui, V. Rao, W. Liao, J. Benner, C. L. Lin and D. Comb, *J. Biol. Chem.*, 1996, **271**, 1732.
27 C. Guan, Y. Liu, Y. Shao, T. Cui, W. Liao, A. Ewel, R. Whitaker and H. Paulus, *J. Biol. Chem.*, 1998, **273**, 9695.
28 J. A. Brannigan, G. Dodson, H. J. Duggleby, P. C. E. Moody, J. L. Smith, D. R. Tomchick and A. G. Murzin, *Nature*, 1995, **278**, 416.
29 Y. Liu, C. Guan and N. N. Aronson, *J. Biol. Chem.*, 1998, **273**, 9688.
30 P. A. Recsei, Q. K. Huynh and E. E. Snell, *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 973.
31 P. Vanderslice, W. C. Copeland and J. D. Robertus, *J. Biol. Chem.*, 1988, **263**, 10583.
32 A. Albert, V. Dhanaraj, U. Genschel, G. Khan, M. K. Ranjee, R. Pulido, B. L. Sibanda, F. v. Delft, M. Witty, T. L. Blundell, A. G. Smith and C. Abell, *Nature Struct. Biol.*, 1998, **5**, 289.
33 X. Duan, F. S. Gimble and F. A. Quiocho, *Cell*, 1997, **89**, 555.
34 K. V. Mills, B. M. Lews, S.-Q. Jiang and H. Paulus, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 3453.
35 B. M. Lew, K. V. Mills and H. Paulus, *J. Biol. Chem.*, 1998, **273**, 15 887.
36 S. Clarke, *Int. J. Peptide Protein Res.*, 1987, **30**, 808.
37 T. Geiger and S. Clarke, *J. Biol. Chem.*, 1987, **262**, 785.
39 C. E. M. Voorter, W. A. de Haard-Hoekman, P. J. M. Van den Oetelaar, H. Bloemendal and W. W. de Jong, *J. Biol. Chem.*, 1988, **263**, 19020.
39 K. Shingledecker, S. Q. Jiang and H. Paulus, *Gene*, 1998, **207**, 187.
40 V. Derbyshire, D. W. Wood, W. Wu, J. T. Dansereau, L. Z. Dalgaard and M. Belfort, *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 11466.

Received, 2nd June 1998
Accepted 17th June 1998