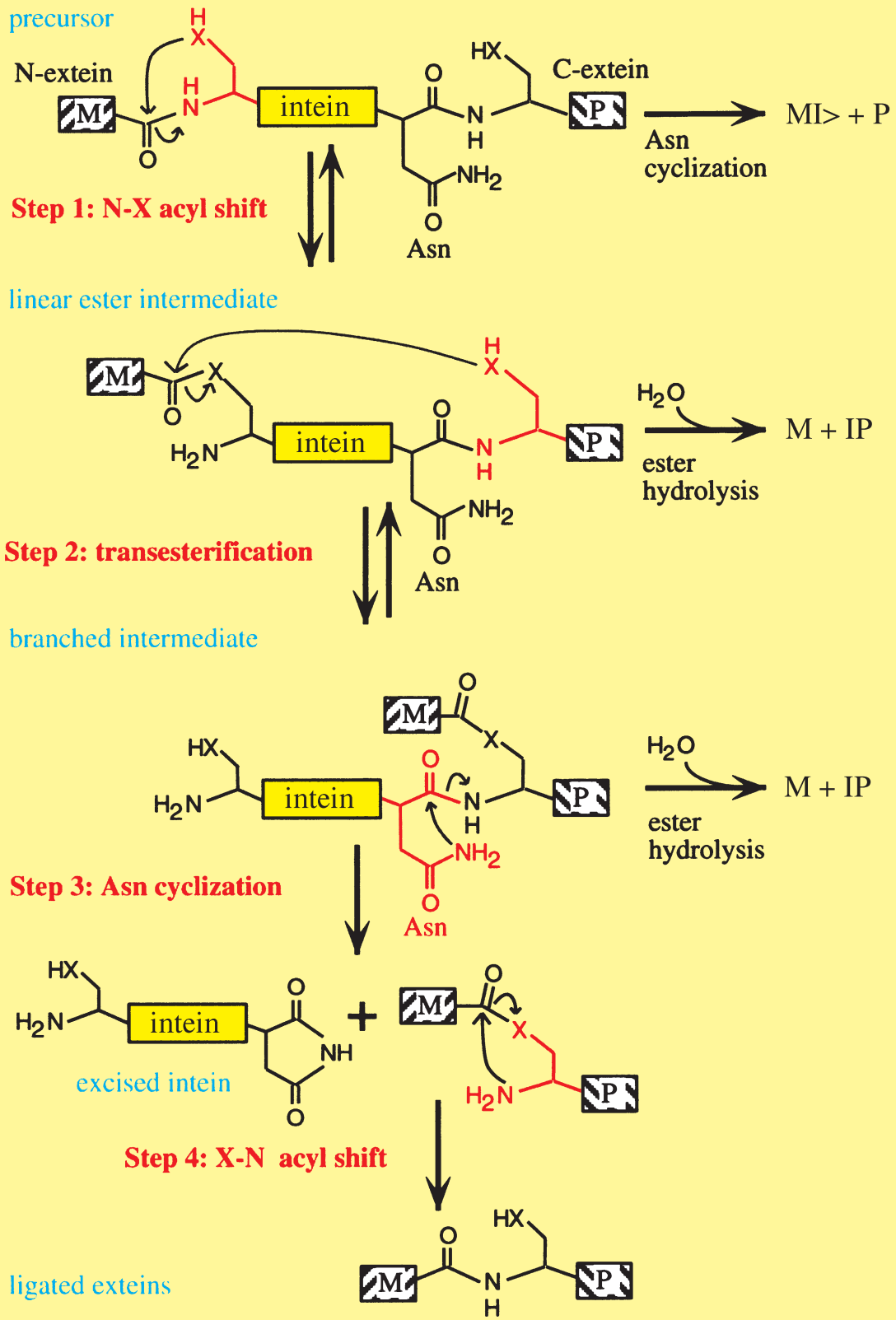
The Currently Accepted Mechanism of Protein Splicing



Dissecting the Chemistry of Protein Splicing and Its Applications

Christopher J. Noren, Jimin Wang, and Francine B. Perler*

In memory of Holger W. Jannasch

Inteins are internal polypeptide sequences that are posttranslationally excised from a protein precursor by a self-catalyzed protein-splicing reaction. The protein splicing domain (≈150 amino acids) activates cleavage of the peptide bonds at the N- and C-terminal splice junctions, with concomitant formation of a new peptide bond between the flanking polypeptides (exteins). The intein plus a single C-extein amino acid can break two peptide bonds, form a new peptide bond, cleave DNA, and initiate mobility of the intein gene. Nearly 100 inteins have been identified with Ser or Cys at the intein N terminus and the triad, His-(Asn/Gln)-(Ser/Thr/Cys), at the C-terminal splice junction. These conserved residues are known to participate in well-studied chemical reactions in other enzymes, which led to a flurry of proposed splicing mechanisms. This review focuses on the three-year period during which a combination of experimental approaches revealed the steps of the protein-splicing pathway, disproving many proposed mechanisms while supporting others. The protein-splicing mechanism involves four coupled nucleophilic displacements: 1) an N→O(S) acyl shift of Ser/Cys at the intein N terminus; 2) a transesterification reaction to form a branched intermediate with two N termini; 3) cyclization of the intein C-terminal Asn/Gln to release the intein; and 4) an O(S)→N acyl shift of Ser/Thr/Cys to form a native peptide bond between the exteins. How the intein facilitates these reactions at the molecular level is only beginning to be elucidated. Understanding the mechanism of protein splicing has resulted in the development of a variety of intein-mediated protein-engineering applications, such as protein purification, addition of fluorescent biosensors, expression of cytotoxic proteins, protein semisynthesis, and segmental labeling of proteins for NMR analysis. Inteins in pathogenic microorganisms provide new targets for drug discovery.

Keywords: autocatalysis · bioorganic chemistry · peptides · protein splicing · transesterifications

1. Introduction

Protein biosynthesis was initially thought to be a simple process in which the genetic information in DNA was directly copied into messenger RNA (transcription), which in turn directed the biosynthesis of protein (translation). There are, however, many genes that encode intervening or terminal amino acids (aa) that are absent in the mature protein. When these sequences are removed from RNA prior to translation of the mature messenger RNA, the process is called RNA splicing and the intervening sequences are termed introns (Scheme 1B). Posttranslational removal of polypeptide segments can occur by protein splicing, autoprocessing, or proteolytic cleavage. In protein splicing, a specific type of intervening sequence, termed an intein,[1] is excised from an internal site in a precursor protein and the surrounding polypeptides (the exteins) are ligated to form the mature protein (Scheme 1A). Protein splicing yields two stable protein products from a single gene, the excised intein and the ligated exteins, and is directly analogous to RNA splicing of introns.

Inteins are members of a growing family of autocatalytic enzyme-like proteins that perform chemical reactions using the same strategies as enzymes, but do not act on multiple substrates. Other examples of autoprocessing include autocleavage of N-terminal nucleophile amidohydrolase precursors (glycosylasparaginase, penicillin acylase, proteasome, etc.), formation of pyruvoyl enzymes, processing of Hedgehog embryonic signaling proteins, RecA-assisted autocleavage of LexA or lambda repressor, and activation of plasma proteins.[2, 3] All of these processes, including protein splicing, are initiated by a nucleophilic attack by the activated side-chain

[*] Dr. F. B. Perler, Dr. C. J. Noren
New England BioLabs, Inc.
32 Tozer Road, Beverly, MA 01915 (USA)
Fax: (+1) 978-921-1350
E-mail: perler@neb.com
Dr. J. Wang
Department of Molecular Biophysics & Biochemistry
Bass Center for Structural Biology, Yale University
New Haven, CT (USA)

  1433-7851/00/3903-0451 $ 17.50+.50/0

thiol or hydroxyl group of Cys, Ser, or Thr at an amide bond. Protein splicing is more complicated than these other autocatalytic reactions since it requires rapid sequential and spatial coupling of four distinct nucleophilic displacements.

Protein splicing proceeds by an intramolecular pathway requiring no exogenous cofactors.[4, 5] All of the information necessary for splicing resides within the intein polypeptide plus the first extein amino acid at the C-terminal splice junction.[4–10] A given intein can excise from a variety of flanking peptide sequence contexts provided that it precedes its cognate C-extein residue, albeit with varying efficiencies.[4–10] To avoid confusion,[11] the term "protein-splicing element" has been used when referring to the splicing-competent intein plus the C-extein Ser, Thr, or Cys residue, as the intein alone cannot mediate protein splicing.

A number of applications have been developed which take advantage of context-independent intein excision, including splicing-dependent protein synthesis,[12, 13] self-cleaving affinity tags for protein purification,[7, 14–16] a novel polypeptide ligation system for protein semisynthesis,[15, 17–20] segmental labeling of proteins for NMR analysis,[12, 20, 21] and addition of fluorescent biosensors.[17, 22] Development of these systems involved genetic and biochemical manipulation of protein-splicing elements, guided by the known mechanism of protein splicing. This review will focus on the stepwise elucidation of the protein-splicing mechanism and modulation of the splicing reaction to create these novel applications. Additionally, we will relate recent structural information to the known mechanistic data in an attempt to dissect the functional requirements for each step in the splicing pathway.

1.1. Inteins Splice as Proteins, Not as RNA

In 1990 two groups independently observed that the open reading frame of the gene for *Saccharomyces cerevisiae* vacuolar ATPase (Sce VMA) was almost twice as large as that of the observed protein product and included 454 aa that had no similarity to other vacuolar ATPases.[23, 24] It seemed unlikely that this intervening sequence was an intron because it had no similarity to known introns, and because it was in-frame with the predicted mature protein—a condition unnecessary for RNA splicing. By 1993 five more in-frame insertions were found in other genes.[4, 25–27] The rapid processing of these native precursors in their natural hosts or in *Escherichia coli* left only indirect methods to prove that intein excision was at the protein level. Modifying the "central dogma" of gene expression demanded direct evidence. Several experimental approaches failed to detect spliced RNA.[10, 23, 24, 27, 28] Silent substitutions that change putative essential intron sequences, but not protein sequence, had no effect on splicing,[26] whereas disrupting the intein reading frame blocked splicing.[10, 26, 27] The breakthrough came in 1993 when scientists at New England BioLabs established the first in vitro splicing system. This allowed purification of an active protein splicing precursor and demonstration of its subsequent splicing in the absence of other proteins or cofactors.[4]


F. B. Perler

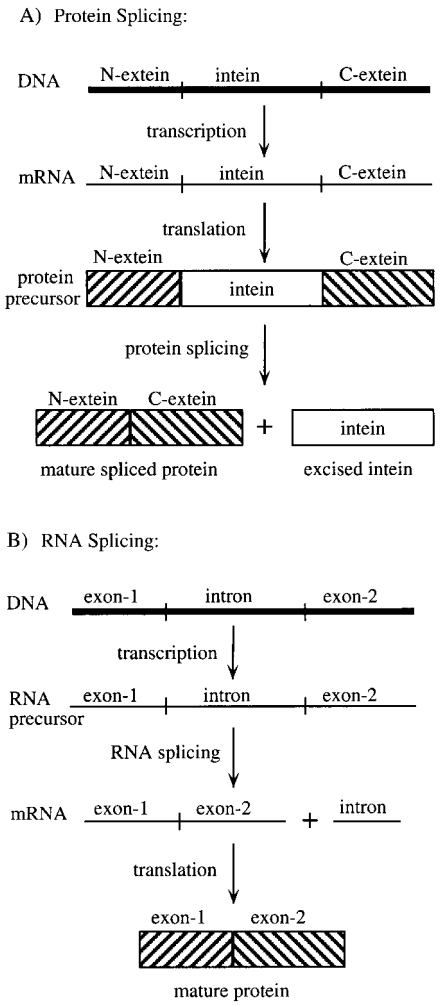

C. J. Noren


J. Wang

Francine B. Perler, born in Brooklyn, New York, received her BS in biology at Brooklyn College, City University of New York in 1972. She earned her PhD in genetics with Clyde A. Hutchison III at the University of North Carolina, Chapel Hill in 1977, cloning ribosomal RNA genes. A postdoctoral fellowship at the Roche Institute of Molecular Biology was followed by a second postdoctoral fellowship in the Biochemistry Department of Harvard Medical School, studying RNA splicing with Dr. Argiris Efstratiadis. Since 1980 she has been a Senior Research Scientist at New England BioLabs, where she studied the molecular biology of filarial parasites and enzymes from extreme thermophiles, especially Vent DNA polymerase. In 1989 she began studying protein splicing, which is now her major focus.

Christopher J. Noren, a native of Manasquan, New Jersey, received his BS in chemistry at the Massachusetts Institute of Technology in 1984. He received his PhD in chemistry from the University of California, Berkeley in 1990, where he codeveloped the unnatural amino acid mutagenesis method in the laboratory of Peter Schultz. He has been at New England BioLabs since 1990, and has been a Senior Scientist since 1993. His efforts at BioLabs have included work on the initial mechanistic studies of protein splicing, as well as developing a series of innovative phage display library technologies.

Jimin Wang, born in Luqiao, Zhejiang, PR China, received his BS in chemistry at Peking University in PR China in 1982, working on transition metal catalysis. He earned his PhD in chemistry at the University of California, San Diego with Professor Joseph Kraut on crystallographic studies of cytochrome C peroxidase in 1988. He then moved to Yale University to join the laboratory of Professor Thomas A. Steitz at the Department of Molecular Biophysics and Biochemistry, where he is a Research Scientist. His research interests are biochemistry and structural biology in DNA replication, transcription, translation, protein folding, and protein degradation.



Scheme 1. A comparison of protein splicing of inteins (A) and RNA splicing of introns (B).

1.2. Intein Mobility and Homing Endonuclease Activity

Similar intein alleles have been found in the same position in homologous genes from closely related and highly divergent organisms,[8, 29–32, 34] although intein distribution is sporadic amongst species. The DNA G + C content and codon usage indicate that some intein genes have recently been transmitted between organisms,[33, 34] and mobility of an intein gene has been experimentally demonstrated in *S. cerevisiae*.[37]

Several types of introns are also mobile genetic elements. An intron encoded homing endonuclease[38] initiates mobility by making a double-stranded break which starts a gene conversion pathway.[13, 30, 35] Homing endonucleases are a class of sequence-specific double-stranded endonucleases that recognize and cleave DNA at or near the intron or intein insertion site in homologous genes lacking the element.[35, 38] Many inteins are bifunctional proteins that also exhibit homing endonuclease activity (see Section 1.3).[26, 28, 37, 38, 41, 43, 44] If an intein-containing allele enters a cell by sexual reproduction, infection, conjugation, transduction, etc., it expresses the intein. The intein endonuclease activity cleaves the intein-minus copy of the gene at or near the intein insertion site ("home"). Since the homing endonuclease recognition site is interrupted by the intein, the intein-plus gene is the only intact copy of the allele available to repair the cleaved intein-minus gene. Repair of the cleaved intein-minus allele using the intact intein-plus allele as the template results in unidirectional gene conversion into an intein-containing allele. In a similar fashion, intein-minus homologues present on phage, viruses, plasmids or other episomal elements can pick up an intein when they pass through a cell containing an intein-plus homologue.

There are four families of homing endonucleases based on the conserved signature motifs that give each family its name.[38] Sequence comparisons indicate that most inteins harbor dodecapeptide motif (DOD) family homing endonucleases, although one intein carries an H-N-H family homing endonuclease (Figure 1 and Section 1.3).[1, 26, 29–31, 35–41] The intein DOD homing endonuclease region is identified by the presence of characteristic motifs, termed intein blocks C, D, E, and H.[29–32, 38, 40] Several lines of evidence suggest that homing endonuclease genes invaded both intron and intein genes during evolution.

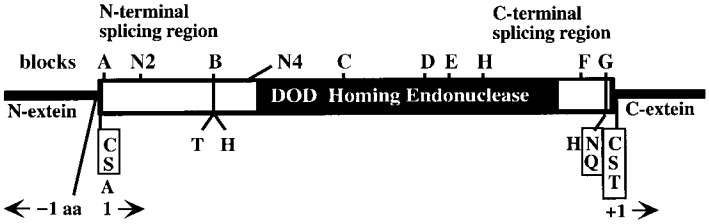


Figure 1. Conserved intein characteristics. A protein-splicing precursor containing an intein with a DOD-type homing endonuclease is depicted with conserved motifs listed above and conserved residues involved in catalysis shown below. Nucleophiles are boxed. Note that some inteins begin with Ala. Intein splicing regions (white boxes) are separated by either an endonuclease domain as depicted or a short linker. By convention, the numbering of amino acids in protein-splicing precursors is as follows: Intein amino acids are numbered from the intein N terminus to its C terminus starting with 1. N-extein amino acids are denoted as negative numbers starting with −1 at the N-terminal splice junction and counting back towards the N terminus of the precursor. The amino acids in the C-extein are numbered from its N terminus (the C-terminal splice junction) to the C terminus of the precursor, beginning with +1 and including a plus sign to denote the C-extein. Arrows indicate the direction of sequential numbering of residues.

1.3. Intein Distribution and Conserved Characteristics

Protein splicing is ubiquitous in nature, with 97 inteins discovered as of July 4, 1999 in unicellular organisms from all three domains of life (eubacteria, archaea, and eucarya) as well as in a virus and a prophage.[29–32, 40] For more up to date information about inteins, see InBase, the intein registry at http://www.neb.com/neb/inteins.html.[29]

A protein is considered to be interrupted by an intein if:
1) extein homologues lack the intervening sequence,
2) the intervening sequence is over 100 aa,
3) the intervening sequence contains intein motifs,
4) the observed protein product is the same size as the predicted open reading frame minus the intein.

The numbering of amino acids in protein splicing precursors is shown in Figure 1.

Inteins range in size from 134 to 600 aa.[29] Some of the larger inteins are bifunctional proteins with endonuclease activity. Ten conserved sequence motifs are present in most inteins (Figure 1).[29–31, 40] In a typical protein splicing precursor, the central intein region (encoding a linker or a homing endonuclease domain) is flanked by the N-terminal splicing region (blocks A, N2, B, N4) and the C-terminal splicing region (blocks F and G), which are in turn flanked, respectively, by the N-extein and the C-extein (block G).[42] Thus, the precursor has a organization like that of Russian nesting dolls (Figure 1). In the Sce VMA intein, part of the homing endonuclease has been transposed into the N-terminal splicing region;[43–45] comparative sequence analysis suggests that this organization may be rare among known inteins.[31, 32, 39, 42] Approximately 10% of inteins have a small linker in place of the homing endonuclease domain.[29]

The intein splicing domain includes blocks A, N2, B, N4, F, and G (Figure 1). No single amino acid is conserved in all inteins. Instead, intein blocks contain related amino acids even in the most conserved positions. All known inteins have a Ser, Thr, or Cys residue on the C-terminal side of both splice junctions, except for two intein families which have Ala at the N-terminal splice junction (Figure 1). N-terminal Edman degradation revealed that the free Tli Pol-2 intein begins with Ser, indicating that the conserved Ser or Cys at the upstream splice junction is the first residue of the intein.[26] All known inteins end in Asn, except two that end in Gln. Block B contains a Thr-(2 aa)-His motif which is also found in serine proteases, where the His is part of the serine protease catalytic triad.[46] A second conserved His is present in block G as the penultimate residue in about 90% of inteins.[29] There is no consensus residue preceding the intein at the C terminus of the N-extein. Since related, but not identical, amino acids are present in conserved motifs, different amino acids act as nucleophiles and assisting residues in each intein. However, conservative substitution of splice junction Ser, Thr, or Cys residues often impairs or inhibits splicing. Therefore, subtle differences in pK_a, nucleophilicity, and conformational properties amongst various inteins potentially increase the complexity of any unified protein-splicing mechanism.

Almost all combinations of splice-junction Ser, Thr, and Cys residues have been observed, except for the presence of Thr at the intein N terminus.[29] This absence of Thr1 may be due to sample size, since mutation of the Tli Pol-2 intein N-terminal Ser1 to Thr permitted efficient splicing.[28] If the multiple intein alleles are counted as one intein family, there is little preference for thiol–thiol (36%), hydroxyl–hydroxyl (18%), or thiol–hydroxyl combinations (46%). However, with the presently available set of 39 different intein alleles with Ser, Thr, or Cys at both splice junctions, there is a bias for Cys at the N-terminal splice junction (82%) and against Cys at the C-terminal splice junction (36%).

1.4. Inteins and Hedgehog Protein Autoprocessing Domains

Although protein splicing has yet to be observed in a multicellular organism, the Hedgehog family of embryonic signaling proteins has an autoprocessing domain which most likely shares a common ancestor with inteins.[42, 45, 48] The Hedgehog autoprocessing domain and the splicing domain of inteins have a common fold (the HINT module),[45] share two motifs (blocks A and B),[46, 47] and perform similar chemical reactions (see Section 2.6).[42, 45, 46, 48] The Hedgehog precursor is cleaved between the N-terminal signaling domain and the C-terminal processing domain. At the same time, cholesterol is attached to the C terminus of the signaling domain, anchoring it to the cell surface where it mediates pattern formation during embryogenesis. Similar autoprocessing domains have been found in several different nematode genes of unknown function.[45, 48, 49]

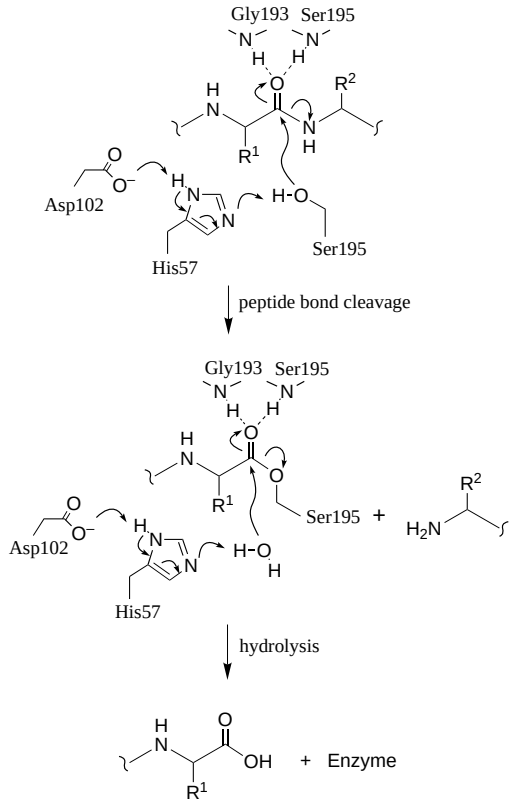
2. Elucidation of the Mechanism of Protein Splicing

Protein splicing results in formation of a native peptide bond in the ligated exteins.[9] The splicing process can be broken down into three distinct, but coupled chemical reactions: two protein cleavage events followed by a peptide ligation event. Many protein-splicing mechanisms have been proposed, based on the known chemical behavior of the conserved residues at the splice junction.

2.1. Similarity with Serine and Cysteine Proteases

In the serine and cysteine classes of proteases,[50] the nucleophilic hydroxyl or thiol side chain of the respective active-site serine or cysteine residue attacks the peptide bond of the substrate at the main chain carbonyl carbon (Scheme 2). This results in cleavage of the scissile peptide bond, with release of the C-terminal cleavage product and formation of an acyl-enzyme intermediate comprising the N-terminal product covalently attached to the enzyme as a (thio)ester. Hydrolysis of the (thio)ester linkage releases the N-terminal polypeptide and regenerates the active protease. The protease reaction is driven to completion by hydrolysis of the acyl-enzyme intermediate. As a result, serine and cysteine proteases can be converted into peptide ligases if water is excluded from the reaction[51] or if the active-site nucleophile is mutagenized into a poor leaving group in the acyl-enzyme hydrolytic step.[52, 53]

Catalysis of the initial nucleophilic attack on the peptide bond is enhanced by increasing both the electrophilicity of the substrate carbonyl group (by interaction of the carbonyl oxygen atom with adjacent positively charged groups, the "oxyanion hole", in the transition state) and the nucleophilicity of the attacking hydroxyl or thiol group.[50] The latter is accomplished by a charge-relay system in which the proton of the attacking species is transferred in the transition state to an adjacent histidine residue, which in turn has its pK_a (and thus basicity) raised by interaction with an adjacent aspartate (in serine proteases) or asparagine residue (in cysteine proteases).[54, 55] This "catalytic triad" of Asp-His-Ser or Asn-His-Cys is the hallmark of all serine or cysteine proteases. The conserved His, Asn/Gln, and Ser/Thr/Cys residues in inteins (Figure 1 and Section 1.3) are reminiscent of protease catalytic triads, which led us to postulate similar functional roles for these residues in protein splicing.[26, 28]



Scheme 2. The serine protease mechanism. The active-site catalytic triad residues (Asp102, His57, and Ser195) of chymotrypsin are shown, along with a hypothetical substrate polypeptide in which the scissile bond is between residue R^1 and residue R^2. Nucleophilic attack by the active-site serine residue is mediated by simultaneous deprotonation by the adjacent His57, which in turn is more basic by interaction with the adjacent negatively charged Asp102. The developing negative charge on the carbonyl oxygen atom is stabilized by interaction with the main-chain protons of Gly193 and Ser195 (the "oxyanion hole"). Tetrahedral intermediates are omitted for clarity.

2.2. An In Vitro Splicing System Leads to the Discovery of a Branched Protein Intermediate

Further elucidation of the splicing mechanism was greatly facilitated by the development of an in vitro splicing system, which allowed time-resolved studies of splicing of a purified precursor protein (Figure 2).[4] This construct consists of the Psp-GBD Pol intein from the Deep Vent DNA polymerase, flanked on the N-terminal side by the *E. coli* maltose binding protein (MBP or M) and on the C-terminal side by a 29-kDa fragment of *Dirofilaria immitis* paramyosin (P). The Psp-GBD Pol intein has serine residues at both splice junctions. This three-part fusion, MIP (for MBP intein paramyosin), facilitated affinity purification of unspliced precursor on amylose resin, as well as rapid identification of splicing intermediates and products by either SDS-PAGE or Western blot. The MIP context was permissive for splicing, but not optimal. Splicing was very slow at 12–20 °C, but more rapid above 30 °C. This temperature sensitivity was not due to the isolation of the Psp-GBD Pol intein from an extreme thermophile, since splicing of the native DNA polymerase precursor in *E. coli* was too fast to isolate precursor at

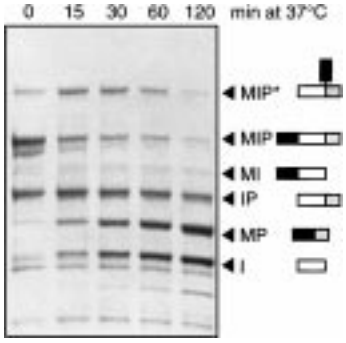


Figure 2. Time-resolved in vitro splicing of purified MIP precursor. MIP, a fusion of MBP, the Psp-GBD Pol intein, and paramyosin, was expressed under nonpermissive splicing conditions and isolated by affinity chromatography on amylose resin.[4] The initial sample contained the MIP precursor (132 kDa), a small amount of a slowly migrating MIP precursor (MIP*), and spliced MP plus free intein (I), and single splice junction cleavage products (M, IP, MI). Time-resolved splicing of the MIP precursor in the absence of added cofactors was observed in SDS-PAGE stained with Coomassie blue. Black box: M (the N-extein); white box: I (the intein); gray box: P (the C-extein).

20 °C. Temperature-dependent splicing due to nonnative exteins turned out to be rather common.[4, 8, 13] Splicing of MIP was pH-dependent, with an optimum at pH 4.5–6.5 and a steady decrease in the rate of splicing up to pH 10, at which point splicing was essentially blocked.[4]

The MIP system allowed identification of a crucial reaction intermediate by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). MIP samples yielded spliced products, MP (MBP paramyosin) and the free intein (I), plus products of single splice junction cleavage (see Section 2.8). N-terminal splice junction cleavage yielded the N-extein plus the intein/C-extein (M + IP), and C-terminal splice junction cleavage yielded the N-extein/intein plus the C-extein (MI + P). In addition to the expected splicing and cleavage products, a slowly migrating species appeared early in the course of the reaction and slowly disappeared as precursor was depleted and product was formed (Figure 2). These kinetics are characteristic of an intermediate and the slow SDS-PAGE migration relative to precursor suggested that the splicing reaction goes through either a covalently multimerized or a branched intermediate. Sequencing of the slowly migrating species by N-terminal Edman degradation revealed two overlapping sequences corresponding to the N termini of the N-extein (MBP) and the intein.[4] Formation and stability of the branched intermediate were pH-dependent: The branched intermediate reverted to precursor at pH 10 and 37 °C, implying that the branch linkage is an ester, since the ester–amide equilibrium would favor the amide at high pH.[56–58] The existence of this branched intermediate implies that the precursor undergoes a rearrangement in which cleavage of the N-terminal splice junction is coupled to acyl migration of the N-extein to the side chain of a residue in the vicinity of the C-terminal splice junction. This branched intermediate structure ruled out a mechanism which predicted that splicing was initiated by Asn cyclization.[59]

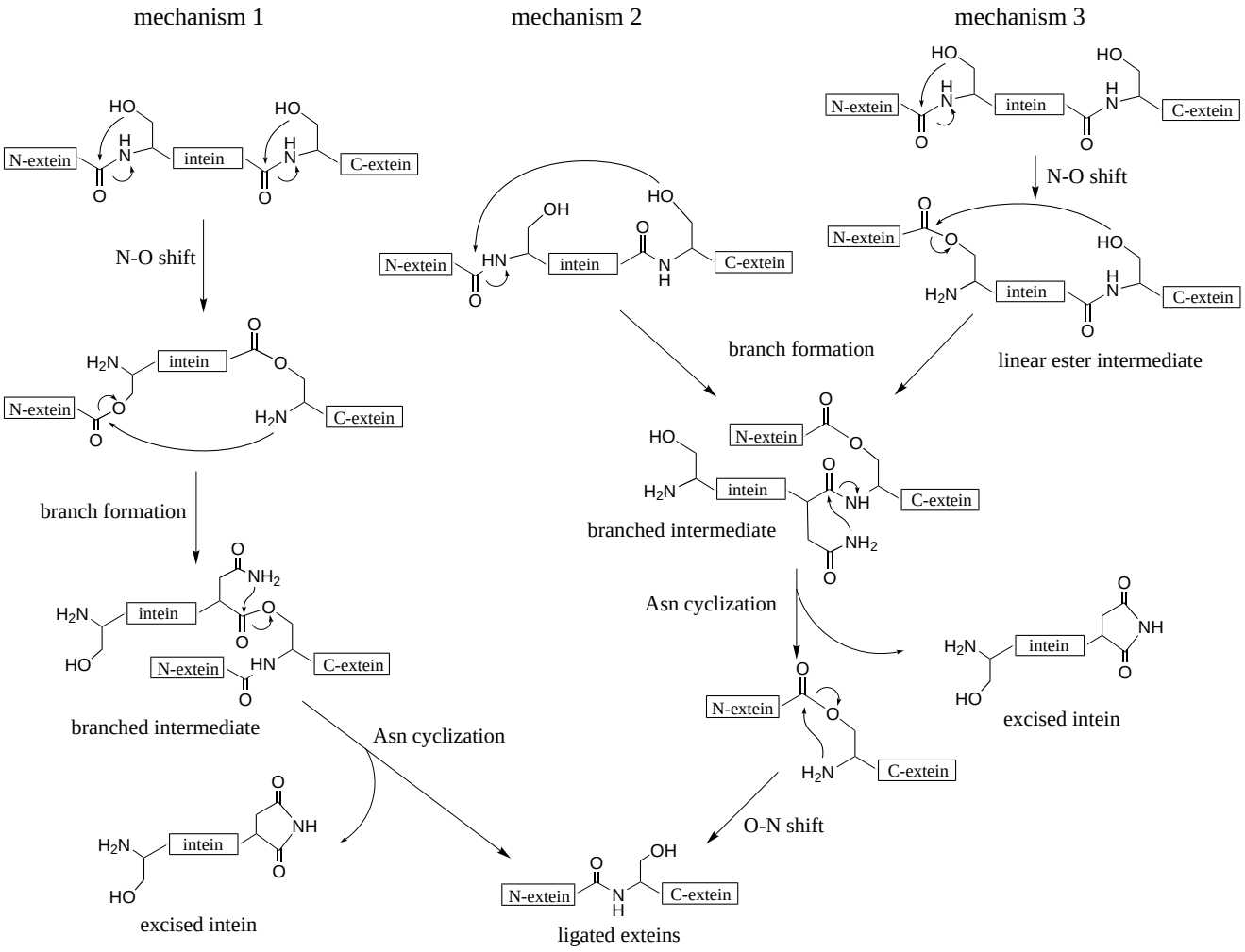
Mutagenesis of the Ser, Thr, or Cys nucleophile at either splice junction to Ala abolished branch formation, whereas mutagenesis of the intein C-terminal His or Asn had little

effect on branch formation.[60] This result immediately ruled out a hypothetical mechanism in which the C-terminal Asn residue is the attacking nucleophile in branch formation.[61]

2.3. Three Proposed Mechanisms Based on a Branched Protein Intermediate

Three more mechanisms were proposed to explain formation and resolution of the branched intermediate in the context of the known mutagenesis data. In mechanism 1 (Scheme 3), independently proposed by Paulus (personal communication in ref. [4]) and Wallace,[62] both splice junctions undergo an N→O(S) acyl migration in which the nucleophilic hydroxyl or thiol group of the conserved Ser/Thr/Cys residues at both splice junctions attack the backbone carbonyl group of the preceding residue, resulting in chemically labile (thio)ester linkages at both ends of the intein. The downstream N→O(S) acyl shift produces a free α-amino group on the first residue of the C-extein, which in turn attacks the labile (thio)ester linkage at the N-terminal splice junction. This reaction breaks the connectivity between the intein and N-extein, and converts the linear (thio)ester intermediate into a branched (thio)ester intermediate in which the intein is joined by an ester linkage to the ligated exteins at the C-terminal splice junction by a peptide bond. The branched intermediate is then resolved by cyclization of the intein C-terminal Asn residue to form a succinimide ring, thereby cleaving the (thio)ester linkage between the intein and the C-extein, liberating both the intein and the ligated extein polypeptides.

In mechanism 2[63] (Scheme 3) the initial acyl migration is distal, rather than proximal. The nucleophilic hydroxyl or thiol side chain of the C-extein + 1 residue attacks the peptide bond at the N-terminal splice junction, which leads directly to a branched intermediate in which the N-extein is joined to the C-extein by a (thio)ester linkage. This reaction is the intramolecular analogue of the initial step in the serine/cysteine protease mechanism, with the branched intermediate equivalent to the protease acyl-enzyme intermediate. In contrast to mechanism 1, the N-extein, rather than the intein, is joined by an ester linkage to the C-terminal splice junction in the branched intermediate. The branched intermediate is re-



Scheme 3. Three proposed protein-splicing mechanisms. All three mechanisms are shown for a hypothetical intein with Ser at both splice junctions, although the overall mechanism is the same if Cys or Thr is substituted at either or both splice junctions. All tetrahedral intermediates and proton transfer steps are omitted for clarity. See text for details.

solved by cyclization of the intein C-terminal Asn residue, releasing the exteins joined by a (thio)ester linkage. In the absence of the stabilizing environment of the intein active site, a spontaneous O(S) →N acyl shift generates the thermodynamically more stable amide linkage between the exteins.[56, 57, 64, 65]

In mechanism 3[63] (Scheme 3) an N →O(S) acyl shift occurs at the N-terminal splice junction, similar to the first step of mechanism 1. The downstream Ser/Thr/Cys nucleophile then attacks the resulting (thio)ester linkage, rather than the amide linkage as proposed in mechanism 2. The remaining steps of mechanisms 2 and 3 are the same.

The characteristics of the branched intermediate were central in eliminating mechanism 1, which predicts that the exteins are joined by an amide linkage and the intein is linked to the C-extein by a (thio)ester bond. In mechanisms 2 and 3, the intein and C-extein are connected in the branched intermediate by an amide and the N-extein by a (thio)ester linkage. Using the MIP system, it was possible to differentiate between these mechanisms by taking advantage of the lability of ester linkages at high pH.[63] The fully denatured MIP branched intermediate decayed to M and IP at 65 °C and pH 9, indicating an alkali-labile ester linkage between the N-extein (M) and the intein/C-extein (IP).[63] This result ruled out mechanism 1, while supporting mechanisms 2 and 3. A similar branched intermediate structure was later shown to exist during splicing of the Sce VMA intein, but was only observed when Cys + 1 was mutated to Ser.[66]

2.4. Resolution of the Branched Intermediate.

In all three mechanisms (Scheme 3), the branched intermediate is resolved by Asn cyclization to cleave the adjacent C-terminal splice junction. Asn cyclization during protein splicing was demonstrated by introducing a methionine five residues upstream from the C-terminal splice junction of MIP, and treating the purified excised intein with the methionine-specific cleavage reagent cyanogen bromide to release the intein C-terminal tetrapeptide.[63] HPLC revealed approximately equal amounts of two peptides, both with the expected composition for the intein C-terminal tetrapeptide. Mass spectrometry showed that the two peptides differed in molecular mass by 18 (a water equivalent), as expected for the hydrolyzed and intact forms of a peptide with a C-terminal succinimide moiety. The succinimide ring was relatively stable in vivo and in vitro.[63] The presence of a succinimide moiety was confirmed using synthetic model peptides, mass spectroscopy, and a colorimetric assay in the Psp-GBD Pol intein[67] and by mass spectrometry of peptides from the Sce VMA intein.[66]

Analogous peptide bond cleavage and formation of a C-terminal succinimide by Asn cyclization occurs in peptides at pH 7.4[68] and during the aging of proteins, such as bovine α-crystallin.[69] Succinimides are also intermediates in the more common spontaneous deamidation and isomerization of internal Asn residues in proteins and peptides.[68, 70–73] Unlike protein splicing and α-crystallin aging, the succinimide in these latter reactions is formed by attack of the main-chain amide nitrogen atom on the side-chain carbonyl group. Hydrolysis of the resultant succinimide at the side-chain or main-chain carbonyl group then produces an internal aspartate or isoaspartate residue, respectively. In peptides, there is a competition between attack by the side-chain amide nitrogen atom on the main-chain carbonyl group versus attack by the main-chain nitrogen atom on the side-chain carbonyl group; in most peptides the former reaction predominates, but the latter reaction is enhanced if Pro, Ser, or His follow Asn in the peptide.[68, 72] Succinimide formation is thought to be limited by the ability to deprotonate the main-chain nitrogen atom, increasing its nucleophilicity.[72] Asn cyclization is also limited in proteins by conformational constraints, since structural information suggests that succinimide formation requires a dihedral φ angle of -120° and χ angle of 120°.[72] Gln isomerizes to form a six-membered ring structure about ten times more slowly than Asn forms a five-membered ring structure due to entropic effects.[72] Splicing of inteins is orders of magnitude more rapid than cleavage of peptides by Asn or Gln (where the half-life is measured in days, not seconds), suggesting that the splicing element must resemble an enzyme in facilitating this reaction under physiological conditions (see Section 2.1).

2.5. Formation of a Peptide Bond between the Ligated Exteins

Following Asn cyclization and intein excision, the exteins remain joined by a (thio)ester linkage between the carbonyl carbon atom of the N-extein − 1 residue and the hydroxyl or thiol side chain of the C-extein + 1 residue. In the absence of the stabilizing effect of the intein, a spontaneous O(S) →N shift occurs, resulting in a thermodynamically more stable amide linkage between the exteins. Model peptides containing internal oxygen ester or thioester linkages at Ser or Cys residues, respectively, rapidly isomerize to the amide form.[64, 65]

2.6. Activation of the N-Terminal Splice Junction for Cleavage

The final piece of the puzzle concerned the conserved nucleophilic Ser1 or Cys1 at the N-terminal splice junction. Unlike mechanism 3, mechanism 2 (Scheme 3) does not postulate a direct catalytic role for this residue. One possibility is that the hydroxyl or thiol side chain forms an essential hydrogen bond, perhaps to the adjacent carbonyl oxygen atom, which would catalyze the nucleophilic attack by the C-extein + 1 residue by stabilizing the developing negative charge in the transition state for the initial tetrahedral intermediate. Substitution of Ser1 with Cys in the Tli Pol-2 and Psp-GBD Pol inteins reduced the rate of splicing and the amount of spliced product after overnight culture,[28, 60] while substitution of Cys1 with Ser abolished splicing of the Sce VMA intein[9, 66, 74] and substitution of Ser1 or Cys1 with Ala blocked splicing in all three inteins. However, N-terminal

splice junction cleavage still occurred after mutation of all three conserved C-terminal splice-junction residues (His-Asn-Ser) to Ala.[60] These data are more consistent with Ser1 and Cys1 playing a direct nucleophilic role, as in mechanism 3.

Mechanism 3 postulates that Ser1 or Cys1 can undergo an N→O(S) acyl rearrangement; the downstream nucleophile then attacks the resulting (thio)ester linkage, rather than the amide linkage proposed in mechanism 2. Similar N→O(S) acyl rearrangements which lead to cleavage of the resultant (thio)ester bonds are observed in other autocatalytic protein-processing reactions, including pyruvoyl formation, processing of the Hedgehog proteins, and N-terminal nucleophile amidohydrolases.[2, 3] Pyruvoyl formation in prohistidine decarboxylase is initiated by an N→O acyl shift of Ser82. Mutation of Ser82 to Cys greatly diminished the rate of pyruvoyl formation, which reflects the less than optimal match between the nucleophile and the surrounding electronic environment of the enzyme active site.[75] A similar phenomenon was seen when conservative substitutions of intein splice junction Ser, Thr, or Cys residues were made,[6, 9, 28, 60, 66, 74, 76, 77] when the active-site nucleophile of serine or cysteine proteases was changed to cysteine and serine, respectively,[78] or when the autoprocessing active-site Thr of Flavobacterium glycosylasparaginase was mutated to Cys.[79]

Hydroxylamine cleavage was used to confirm the acyl shift in a Ser82Cys mutant of prohistidine decarboxylase.[75] In proteins at neutral pH, internal thioester linkages are much more sensitive to cleavage with hydroxylamine than amide and oxygen ester linkages.[58] The Ser82Cys mutant of prohistidine decarboxylase, unlike the wild-type enzyme, was sensitive to hydroxylamine cleavage. Internal thioesters in proteins can also be displaced by small thiol reagents, resulting in protein cleavage. A similar strategy was used to detect thioester intermediates in processing of the Hedgehog protein and the Flavobacterium glycosylasparaginase.[48, 60, 79, 80] Mutation of the Psp-GBD Pol intein Ser1 to Cys reduced both splicing and branched intermediate formation, but prolonged incubation in the presence of hydroxylamine, free Cys, or thiol reagents resulted in increased levels of N-terminal splice junction cleavage relative to wild-type precursor, diagnostic of the presence of a thioester linkage at the N-terminal splice junction.[60, 80] In the MIP system, hydroxylamine cleavage at Cys1 also occurs in the presence of additional mutations that block branch formation, indicating that the linear (thio)ester precedes the branched (thio)ester in the splicing pathway.[60] The position of hydroxylamine cleavage was confirmed by mass spectrometry and HPLC analysis of a C-terminal cyanogen bromide fragment derived from the cleaved N-extein with comparison to a model peptide containing the expected C-terminal hydroxamate moiety.[80]. Similar results were later obtained with the Sce VMA intein.[66]

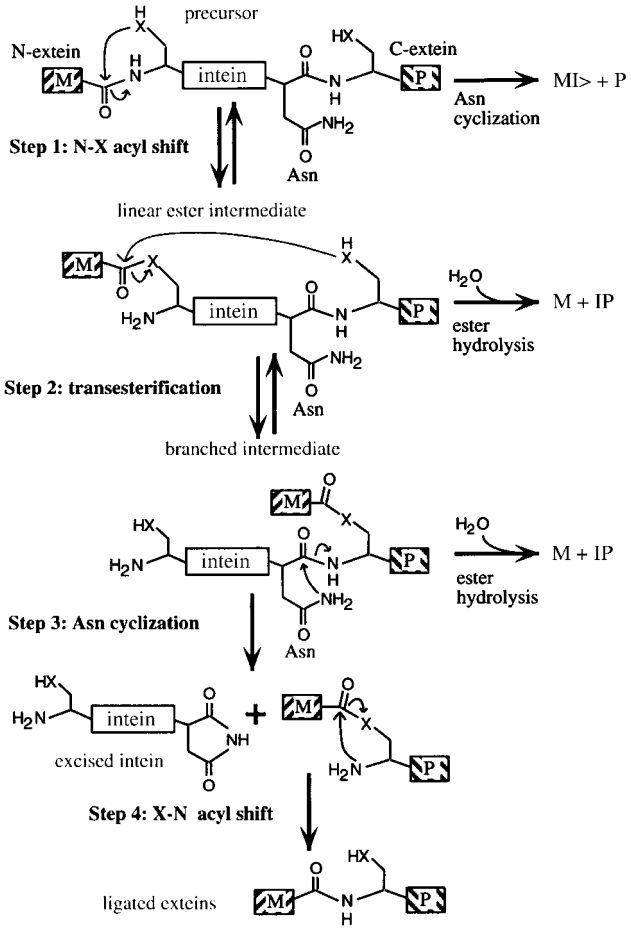
Taken together, these data strongly implicate an N→O(S) acyl shift at the upstream splice junction as the initial step of protein splicing according to mechanism 3. However, two families of putative inteins have an N-terminal Ala residue (see Section 1.3). Preliminary evidence indicates that at least one of these inteins is capable of splicing.[81] We propose that these Ala1-containing inteins have evolved to splice using mechanism 2, since all other intein motifs are intact.

2.7. The Protein-Splicing Mechanism

The widely accepted mechanism for protein splicing of most inteins is shown in Scheme 4 and consists of four steps:

1) attack by the side chain of the first intein residue (Ser, Thr, or Cys) on the preceding carbonyl group, resulting in an acyl shift of the N-extein to the side chain of the first intein residue;
2) a transesterification reaction in which the hydroxyl or thiol group on the first C-extein residue (Ser, Thr, or Cys) attacks the (thio)ester linkage, resulting in transfer of the N-extein to the C-terminal splice junction;
3) cleavage of the amide linkage at the intein C terminus by Asn or Gln cyclization to release the free intein; and
4) spontaneous rearrangement of the (thio)ester linkage between the ligated exteins to the more stable peptide bond amide linkage.

Processing of the Hedgehog protein provides an interesting variation of the intein-mediated transesterification reaction.[48] In the Hedgehog system, nucleophilic displacement



Scheme 4. The currently accepted mechanism of protein splicing and single splice junction cleavage. X represents either the oxygen present in the side chain of Ser or Thr or the sulfur present in Cys. In some inteins, Asn is replaced by Gln. Inteins with Ala1 probably splice using mechanism 2 (Scheme 3). All tetrahedral intermediates and proton transfer steps are omitted for clarity. Points in the pathway where single splice junction cleavage can occur are also indicated. See text for details. "M" represents the N-extein, and "P" the C-extein. The presence of a C-terminal succinimide is indicated by ">".

of the thioester linkage between the signaling and autoprocessing domains occurs when the hydroxyl group of cholesterol attacks the thioester. This results in the covalent linkage of cholesterol to the C terminus of the signaling domain, which serves to anchor the signaling domain on the surface of the cell where it controls important aspects of embryonic development.[48]

2.8. Side Reactions

In addition to the expected excised intein and ligated extein products, protein splicing typically yields products from side reactions, especially when splicing precursors are expressed in heterologous hosts, or when inteins are cloned between foreign exteins. These products arise from cleavage at either or both splice junctions without concomitant ligation and are clear extensions of the mechanism (Scheme 4). Cleavage at the N-terminal splice junction arises from competing hydrolysis of the linear (thio)ester intermediate versus transesterification or by hydrolysis of the (thio)ester in the branched intermediate prior to Asn or Gln cyclization. Cleavage at the C-terminal splice junction arises by Asn or Gln cyclization in the absence of N-extein transfer. Mutating catalytic residues at one splice junction generally increases the cleavage side reaction at the other junction.[6, 9, 14, 60, 66, 74] For example, replacement of the intein N-terminal residue with Ala results exclusively in cleavage at the C-terminal splice junction, while replacement of the intein C-terminal Asn with Ala results in cleavage at the N-terminal junction.

2.9. The Role of Proximal Extein Amino Acids

Splicing of chimeric precursors is usually slower than in the native precursor, with more abortive single splice junction cleavage products.[4–10] In our experience, splicing is successful in less than 25% of chimeric precursors. Moreover, it is often temperature-dependent, suggesting that folding of the intein is impaired.[4, 8, 13] Since all three nucleophiles in the protein-splicing pathway are splice-junction residues, suboptimal proximal extein sequences are likely to inhibit splicing by destabilizing the active site either chemically or by sterically interfering with packing at the active site.

Although the -1 residue at the N-terminal splice junction does not actively participate in the splicing reaction as a nucleophile or assisting group, its carbonyl carbon atom is the target of three of the four nucleophilic displacements. Several studies have shown that N-extein proximal residues can inhibit or block splicing or rescue splicing defective mutants.[6, 8, 14, 76, 77] Two systematic studies have demonstrated that some -1 substitutions can block splicing of the Sce VMA and *Mycobacterium xenopi* gyrase A (Mxe GyrA) inteins, although residues that block splicing of one intein may have negligible effect on the other intein.[6, 14] Amino acids that are most similar to the cognate -1 residue worked best in these studies. Certain amino acids at the -1 position can either reduce the rate of the initial acyl shift or block transesterification.[14] These observations imply that inteins have coevolved with their extein sequences to maximize splicing efficiency and that it is best to choose intein insertion sites that are similar to the native insertion site.

2.10. Facilitators of Catalysis and Intein Structure

As discussed in Section 2.1, serine and cysteine proteases facilitate catalysis using an oxyanion hole to increase the electrophilicity of the substrate carbonyl and a charge relay system to increase the nucleophilicity of the attacking hydroxyl or thiol group.[50] Inteins probably use similar strategies. Mutagenesis studies indicate that the conserved His in block B assists in reactions at the intein N terminus[77] and in the Hedgehog protein,[46, 82] but fail to conclusively establish the role of the intein penultimate His residue of block G.

The penultimate His residue is required for splicing and C-terminal splice-junction cleavage of the Psp-GBD Pol intein.[60] However, the His residue of block G does not appear to be universally required for splicing, although mutation of this residue generally inhibits splicing to varying degrees depending on the intein, the extein, and the substituted amino acid.[6, 9, 60] Moreover, about 10% of inteins naturally have a residue other than His at this position, suggesting that when a mutation naturally occurs, the intein can evolve to splice in the absence of a penultimate His.[29] When four inteins which naturally have Ser, Ala, Phe, or Gly as penultimate residues were tested for their ability to splice in *E. coli*, only the *Chlamydomonas* ClpP intein failed to splice.[81, 83–85] Mutation of the penultimate Gly residue of *Chlamydomonas* ClpP intein to His[83] and of the penultimate Phe residue of *M. jannaschii* phosphoenolpyruvate synthase intein to His[85] significantly improved splicing. On the other hand, mutation of the penultimate Gly residue of *Methanococcus jannaschii* RNA polymerase A′ intein to His significantly inhibited splicing, resulting in accumulation of the branched intermediate.[85] Thus, splicing in *E. coli* of inteins naturally containing penultimate amino acids other than His may improve or worsen when "reverted" to His, depending on the intein. These conflicting results may reflect inteins at different stages of evolution after mutation of their penultimate His residue. Taken together, these data suggest that the intein penultimate His may facilitate splicing and branch resolution by Asn cyclization.

Structural data support the role of both conserved His residues in facilitating protein splicing. The crystal structures of two inteins and part of a Hedgehog autoprocessing domain have been determined.[45, 86, 87] The Sce VMA intein and the N-terminal portion of the *Drosophila* Hedgehog protein autoprocessing domain were both crystallized after excision from the precursor, while the Mxe GyrA intein was crystallized with a single N-extein residue (Ala -1) and Cys1 mutated to Ser. The Cα trace of all three proteins can be superimposed over much of the splicing domain[42, 45, 48, 86, 87] and form a conserved fold termed a HINT (*H*edgehog *int*ein) module.[45] The HINT module consists of two structural subdomains with superimposable Cα traces that are related by a pseudotwofold axis of symmetry and have a single

hydrophobic core.[45] The N and C termini are brought into proximity upon folding of the HINT module, which is composed entirely of β strands and loops.

Unfortunately, the C-extein nucleophile and the scissile bonds are absent from these structures, except for the N-terminal splice junction in the Mxe GyrA intein. In the Mxe GyrA intein structure, Ala $-$ 1 and Ser1 are in a *cis* conformation, which may provide part of the driving force for protein splicing.[87, 88] However, Ala at the $-$1 position inhibits splicing by 90% and decreases the initial rate of thiol-induced N-terminal cleavage fivefold when the Mxe GyrA intein is in the MIP context.[14] The positions of Ala $-$ 1 and the C-terminal Asn in this structure may be different than in a larger precursor due to interaction between the charged N-terminal α-amino group of Ala $-$ 1 and the C-terminal Asn198 carboxylate group (Figure 3).

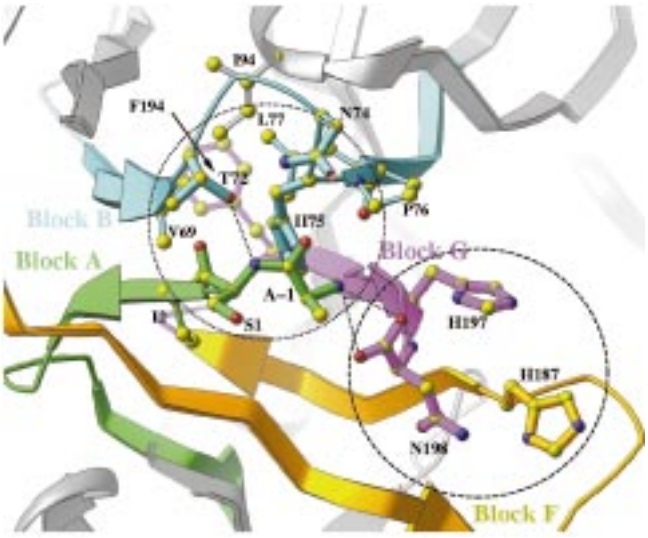


Figure 3. The structure of the Mxe GyrA intein active site. Amino acids present at the Mxe GyrA intein active site and the conserved motifs are depicted based on the Mxe GyrA intein crystal structure.[87] Two potential oxyanion holes are indicated by the dashed circles. A pair of dashed lines indicates a potential hydrogen-bond interaction between the free amino and carboxy groups of the protein, which may distort the position of these groups in this protein relative to their position in a precursor with larger exteins. A third dashed line indicates a potential hydrogen bond between the hydroxyl group of Thr72 in block B and the nitrogen atom of Ser1 at the scissile bond.

The crystal structures nonetheless provide insight into some of the groups assisting in splicing reactions. The structures reveal an intein with a preorganized active site including two potential oxyanion holes. In all three structures, several block B residues are within hydrogen-bonding distance of the N terminus of the HINT module, potentially interacting in the precursor with either the N-terminal carbonyl oxygen atom at the scissile bond or the amine leaving group, forming an oxyanion hole reminiscent of serine and cysteine proteases (see Section 2.1). These include, but are not limited to, the conserved His and Thr residues (Asn in the Sce VMA intein) in block B (Figure 1 and Section 1.3). In the Mxe GyrA intein structure, His75 of block B is within hydrogen-bonding distance to the Ala $-$ 1 carbonyl oxygen atom, and Thr72 and Asn74 of block B are within hydrogen-bonding distance to the main chain nitrogen leaving group (Figure 3).[87]

Several residues could potentially form a second oxyanion hole to facilitate Asn cyclization (Figure 3). In both inteins, the penultimate His residue is positioned to form a hydrogen bond with the C-terminal carboxylate group, and not the Asn side-chain nitrogen nucleophile. These data suggest that the His residue acts by either stabilizing the developing negative charge on the tetrahedral intermediate of the Asn cyclization reaction or by making the main-chain amine a better leaving group, but not by increasing the nucleophilicity of the attacking side-chain nitrogen atom.[45, 86, 87]

No amino acid is interacting with the side-chain hydroxyl or thiol group at the N-terminal splice junction to increase its nucleophilicity.[45, 86, 87] All three structures are of elements with an N-terminal Cys residue. It would be interesting to see the structure of an intein with an N-terminal Ser residue, since such inteins would probably require assistance to increase the nucleophilicity of Ser1, due to its high pK_a.

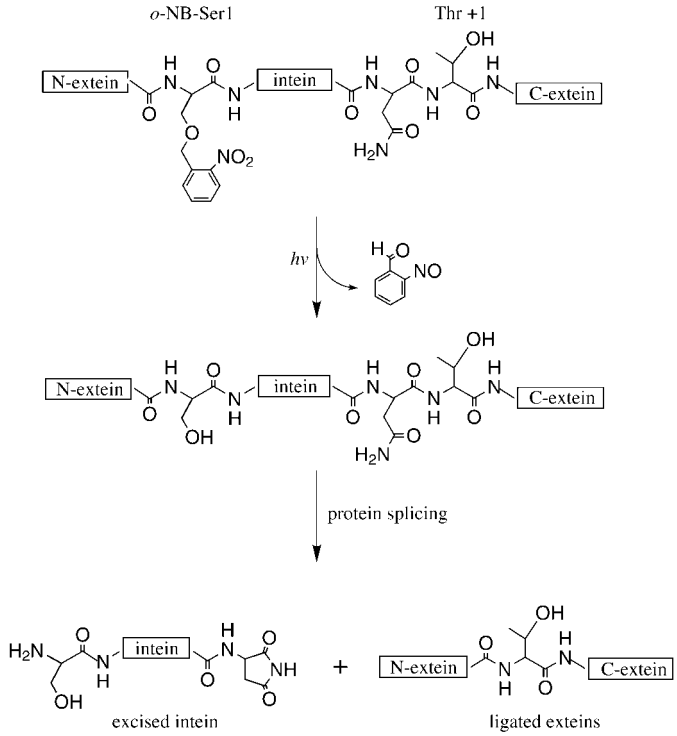
More questions about protein-splicing catalysis remain than have been answered. Which steps in the pathway does the oxyanion hole at the N-terminal splice junction assist, and is there a second oxyanion hole present at the C-terminal splice junction? What type of conformational changes, if any, are required to achieve the four coupled nucleophilic displacements, given that three of them attack the same carbonyl carbon atom? Must side chains move to allow each new nucleophilic group to attack? What residues help align the attacking nucleophile with the electrophilic center? Why do conservative substitutions of Ser, Thr, and Cys usually inhibit splicing? Is the unusual *cis* conformation at the Mxe GyrA intein N-terminal splice junction present in the native precursor, and if so do the resulting steric constraints limit the allowable amino acids at the $-$1 position? A thorough understanding of the protein-splicing mechanism awaits the determination of the structure of precursors and intermediates from inteins with both Cys and Ser splice-junction residues.

3. Applications

3.1. Control of Splicing

Applications involving control of splicing fall into two categories: pharmaceutical "switches" and tools for protein engineering. Inteins are present in essential genes of human pathogens such as *Mycobacterium tuberculosis*, *Mycobacterium leprae*, and *Candida tropicalis*. Since intein-containing precursors are inactive, chemical reagents that block splicing would have significant pharmacological value. Controllable inteins inserted into foreign contexts provide a method for activating an enzyme on command, allowing biosynthesis of cytotoxic proteins or controllable gene knockouts. Splicing was first controlled by insertion of the intein into a suboptimal extein context or by intein mutation, such that precursor biosynthesis and splicing occur at different temperatures.[4, 8, 13, 60]

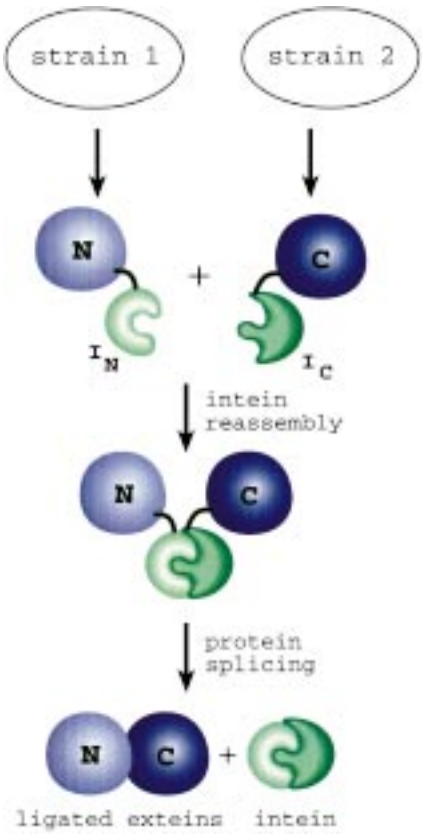
Protein splicing has also been controlled by a novel use of photochemistry and site-directed mutagenesis (Scheme 5).[89] With a chemically aminoacylated suppressor tRNA and an in



Scheme 5. Control of splicing using photoactive groups. Protein splicing of the Tli Pol-2 intein with a caged photoreactive group is initiated by photolysis of the 2-nitrobenzyl ether of Ser1. *o*-NB denotes the *o*-nitrobenzyl protecting group.[89]

vitro translation system,[90] *O*-(2-nitrobenzyl)serine was incorporated in place of Ser1 in the Tli Pol-2 intein. The "caged" protecting group can be released from the precursor by brief irradiation with visible light, initiating splicing by restoring Ser1. Photochemically activated protein splicing should allow a range of time-resolved studies relevant to the mechanism of protein splicing.

Trans-splicing in vitro is another method of controlling splicing by splitting the precursor gene into two fragments which are separately expressed in different cultures (Scheme 6).[12, 21, 91, 92] After purification, the precursor fragments are mixed under native or denaturing conditions, depending on solubility and reassociation requirements. Reassembly activates the splicing element, resulting in ligation of the exteins and release of the intein fragments. In all cases to date, the precursor has been split in the dispensable endonuclease domain of the intein. Two out of three split sites led to functional trans-splicing in the Psp-GBD Pol intein, and the released split intein complex retained homing endonuclease activity.[12] Trans-splicing allows the expression of extremely cytotoxic proteins as well as segmental protein modification or labeling. Trans-splicing may also provide a new tool to study protein folding. In another example of nature mimicking biotechnology (or vice versa), the catalytic subunit of the replicative DNA polymerase of *Synechocystis* sp. has been split within its intein by a



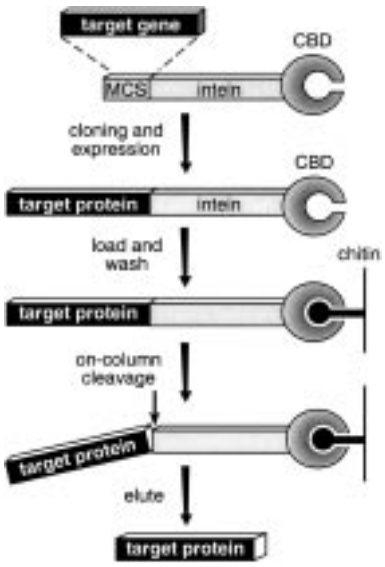
Scheme 6. Trans-splicing: control of splicing by precursor fragmentation. The precursor gene can be split within the intein and each fragment expressed in separate cultures. Reassembly occurs in vitro. The intein fragments ($I_N + I_C$) reconstitute an active splicing element that then mediates ligation of the extein fragments (N, C).

genome inversion, but trans-splicing in vivo restores DNA polymerase activity to maintain viability of this organism.[84]

3.2. Protein Purification Vectors

Understanding the mechanism of protein splicing has allowed the conversion of an intein into the more common type of autocleavage element, an autoprotease that cuts a single adjacent peptide bond. Cleavage at only the N-terminal splice junction is achieved by mutating one or more C-terminal splice-junction residues to prevent Asn cyclization while allowing thioester formation.[14, 16, 60] The thioester linkage can then be cleaved with thiol-specific reagents, including free cysteine, or by nitrogen nucleophiles such as hydroxylamine. Several methods have been used to specifically activate the intein C-terminal splice junction, including mutation of the Sce VMA intein penultimate His[7] and mutation of the -1 aa in a temperature-controllable Mxe GyrA intein.[14]

Converting a modified intein, which can only cleave one splice junction, into a purification vector is achieved by 1) introducing a multiple cloning site adjacent to the active splice junction for cloning of the target gene and 2) cloning the gene for an affinity purification tag within the intein or adjacent to the inactive splice junction (Scheme 7). Precursors can be expressed in any eucaryotic or procaryotic host. Intein cleavage is directly activated on the affinity purification



Scheme 7. Intein purification vectors. Purification of target proteins can be achieved by cloning the target protein in-frame with the single active N- or C-terminal splice junction of the modified intein using the multiple cloning site (MCS).[7, 14–16] After expression under nonpermissive autocleavage conditions, the fusion protein is isolated on a chitin affinity column using the chitin binding domain (CBD) affinity tag that is adjacent to the inactive splice junction. Cleavage is then induced on the column by various methods including thiol reagents or temperature shift, followed by elution of the free target protein.

column, resulting in elution of the target protein while the remainder of the fusion is retained. Two purification vectors are commercially available (New England BioLabs, Inc.) in which the C terminus of the target protein is fused to the active N terminus of the intein, and a 5 kDa bacterial chitin binding domain affinity tag is fused to the inactive C terminus of the intein. The IMPACT system (intein-mediated purification with an affinity chitin-binding tag) employs a modified 454 aa Sce VMA intein,[16] while a second vector employs a modified 198 aa Mxe GyrA intein.[14, 15] In a third commercially available purification vector which is also part of the IMPACT system (New England BioLabs, Inc.), a multiple cloning site for the target protein is adjacent to the C terminus of a modified Sce VMA intein in which the central endonuclease domain has been replaced by the 5-kDa chitin binding domain and ten codons are in-frame with the N-terminal splice junction.[7] In this mutated intein, Asn cyclization is activated by thiol-induced cleavage of the peptide at the intein N terminus.

Intein-mediated protein purification has several advantages over other purification methods. No expensive protease is necessary, and cleavage occurs only at the splice junction and not at protease-sensitive sites in the target protein. Furthermore, purification of a target protein free of vector-derived amino acids can be accomplished in one step, since purification and cleavage occur on the same column.

3.3. Intein-Mediated Polypeptide Ligation and Related Techniques

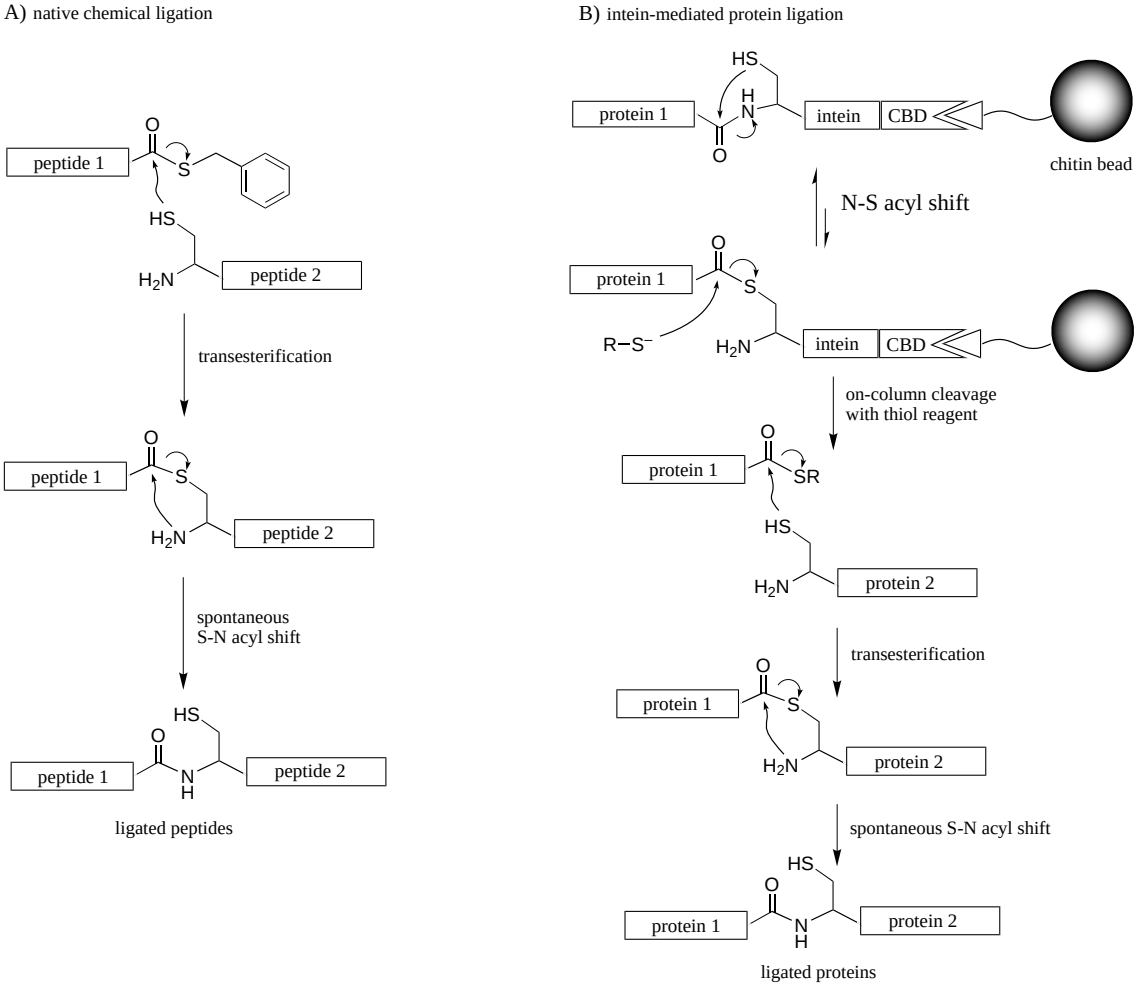
1998 saw the convergence of two previously unrelated fields, protein splicing and protein semisynthesis, resulting in generation of a new protein engineering technique termed either "expressed protein ligation" (EPL)[17, 19, 20, 22] or "intein-mediated protein ligation" (IPL).[15, 18] The elucidation of the protein-splicing mechanism occurred simultaneously with the development of a procedure for protein semisynthesis, termed "native chemical ligation".[65] In native chemical ligation, a synthetic peptide with a C-terminal α-thioester moiety is mixed with a second peptide or protein with an N-terminal cysteine (Scheme 8A). A highly efficient and chemoselective intermolecular reaction takes place, forming a thioester-linked intermediate which spontaneously rearranges by an S $\rightarrow$N acyl rearrangement to form a peptide bond between the ligated peptides. Thus, native chemical ligation follows all the steps in the protein-splicing pathway except for the Asn cyclization reaction which releases the intein from the ligated exteins (Scheme 4). Native chemical ligation is limited by the ability to make large peptides by solid-phase peptide synthesis and the inability to express a polypeptide with a C-terminal α-thioester or to selectively generate a C-terminal α-thioester in an unprotected polypeptide.

IMPACT and similar intein vectors (see Section 3.2) provide unique tools for biosynthesis of an α-thioester at the C terminus of a target protein, which can then be utilized in various types of chemoselective condensation reactions, including modified native chemical ligation. Other intein vectors yield a polypeptide with an N-terminal Cys residue after cleavage at the intein C-terminal splice junction.[14, 15] Alternatively, protease digestion can generate an N-terminal Cys residue in a biologically synthesized polypeptide.[20, 22] To date, the Mxe GyrA intein,[14, 18] the Sce VMA intein,[16–20, 22] and the *Methanobacterium thermoautotrophicum* ribonucleoside – diphosphate reductase (Mth RIR1) intein[15] have been used to express polypeptides with a C-terminal α-thioester.

Initial studies involved condensation of synthetic peptides with intein-derived, biologically expressed proteins containing an α-thioester between the C terminus of the target protein and the N terminus of the intein (Scheme 8B) to express cytotoxic enzymes,[18] to test the effect of phosphorylation on kinase activity and conformation,[17] and to examine protein – protein interactions.[19] Semisynthesis was achieved by:

1) truncating the target gene immediately prior to the last cysteine codon or at a nonconserved point near the C terminus;
2) expressing the truncated enzyme fused to the N terminus of a modified intein;
3) purifying the truncated enzyme by affinity chromatography; and
4) mixing the truncated enzyme with a synthetic peptide comprising the remainder of the enzyme with an N-terminal Cys, and condensing in the presence of a thiol reagent, such as 2-mercaptoethanesulfonic acid (MESNA) or thiophenol. Alternatively, the thiol reagent can be added first to cleave the truncated protein from the column as an α-thioester, followed by addition of the synthetic peptide (Scheme 8B).

Ligated RNase A required refolding for activity, but the HpaI restriction enzyme did not.[18] The C-terminal Src kinase (Csk) proteins (50 kDa) catalyze the phosphorylation of a



Scheme 8. Comparison of native chemical ligation and intein-mediated protein ligation. A) Native chemical ligation[65] involves the coupling of two synthetic peptides, the N-terminal peptide having a C-terminal thioester linkage and the C-terminal peptide (or protein) having an N-terminal cysteine residue. Mixing results in transesterification to yield a ligation product in which the two peptides are joined by a thioester linkage with an adjacent free α-amine. The thioester linkage spontaneously rearranges to the more stable amide linkage. B) Intein-mediated expressed protein ligation[15, 17–20, 22] utilizes identical coupling chemistry as native chemical ligation, but the α-thioester on the N-terminal segment is generated by on-column cleavage of a recombinant protein fused to a cleavage-competent intein. This circumvents the length requirements imposed by synthetic peptide chemistry. CBD refers to the chitin binding domain, which is used to immobilize the intein, and the triangle represents a chitin moiety attached to a bead. All tetrahedral intermediates and proton transfer steps are omitted for clarity.

conserved tyrosine within the C-terminal tail of Src kinase family members (SH2 domain).[17] Using a similar synthetic approach with an IMPACT vector (Sce VMA intein) and thiophenol, either a phosphorylated or unphosphorylated synthetic peptide was condensed with the truncated human Csk. The resultant kinase was significantly more active when the phosphorylated peptide was ligated, confirming the role of phosphotyrosine for activity. A fluorescent tag was included in each peptide, allowing analysis of the conformation by limiting proteolysis, which indicated that phosphotyrosine induced a change in Csk conformation. In the third example, a C-terminal peptide was ligated to a truncated σ^{70} subunit of *E. coli* RNA polymerase to demonstrate that this region of the protein is involved in protein–protein interaction in the holoenzyme and in promoter recognition.[19]

The second generation of IPL utilizes a biologically expressed polypeptide with an N-terminal Cys in place of a synthetic peptide. The N-terminal Cys residue can be generated with two intein vectors in which the N terminus of the target protein is fused to the intein C terminus,[14, 15] but not with the commercial IMPACT Sce VMA intein C-terminal splice-junction cleavage system[7] (see Section 3.2). Alternatively, an N-terminal Cys residue can be generated by protease digestion of an expressed polypeptide[20, 22] In each example, two large polypeptides were solubly expressed in bacteria and condensed without the need for denaturation or renaturation.[15, 20, 22]

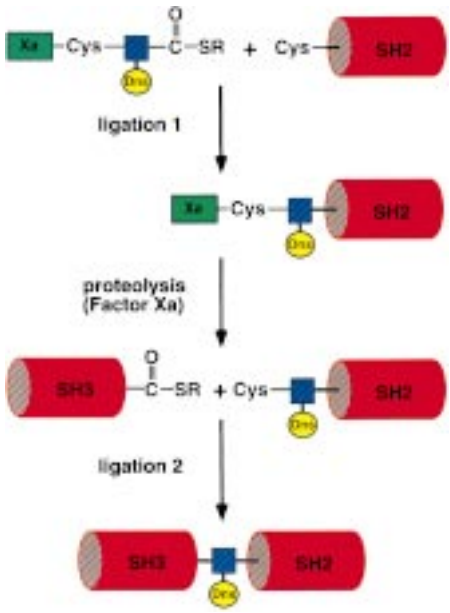
These same techniques can be used to specifically label the N or C terminus of a protein with any type of tag, which can either be incorporated into a peptide or displace the intein from the α-thioester intermediate. For example, the Hedgehog autoprocessing domain generates a similar thioester linkage which allows attachment of cholesterol to the C terminus of the released signaling domain by a transesterification reaction involving the hydroxyl group of cholesterol.[46, 48]

A major breakthrough in NMR spectroscopy for protein structure determination was achieved with IPL.[20] NMR

analysis is limited by the loss of resolution with increasing number of amino acids. Segmentally labeling a protein reduces the spectral complexity and has been used to study large proteins whose solution structures were previously unsolvable by NMR spectroscopy. With IPL, a segment or domain of a large protein is isotopically labeled in bacteria with ^{13}C or ^{15}N and then ligated to the remaining unlabeled segment or domain of the protein, also expressed in bacteria. In a series of separate experiments, different parts of the protein are labeled prior to ligation by IPL, allowing the structure of the labeled region to be determined in the context of the entire folded protein. This technique also provides a method of studying the effect of ligand binding or domain interactions by chemical shift perturbation to elucidate structure – activity relationships by NMR spectroscopy. Similar strategies of segmental labeling are possible by trans-splicing[12] (Section 3.1) and NMR analysis of a small polypeptide has been reported after trans-splicing.[21] Although both trans-splicing and IPL achieve the same result, trans-splicing is limited by the difficulties of identifying appropriate sites for intein insertion, the need to reconstitute an active intein, and the potential need to denature the precursor fragments prior to reassembly (see Section 3.1).

In another elegant set of experiments, Muir and co-workers have used intein-mediated expressed protein ligation to introduce a protein biosensor between the Src homology 3 (SH3) and Src homology 2 domains (SH2) of the Abelson nonreceptor protein tyrosine kinase (Abl).[22] A synthetic peptide containing an environmentally sensitive 5-(dimethylamino)naphthalene-1-sulfonamide (Dns) fluorophore was sequentially ligated to the Abl-SH2 domain at its C terminus and the Abl-SH3 domain at its N terminus using IPL and native chemical ligation strategies (Scheme 9). The resultant semisynthetic protein has an environmentally sensitive dansyl group in the linker between the SH2 and SH3 domains which was able to report conformational changes upon ligand binding. This protein engineering strategy is generally applicable to the insertion of any type of tag, fluorophore, or unnatural amino acid into a protein of any size or sequence without the limitations of methods involving chemically misacylated suppressor tRNA's[90] or solid-phase peptide synthesis.

Recently, inteins have also been used to generate cyclized proteins[93] and proteins with a C-terminal thiocarboxylate for protein semisynthesis.[94] The C-terminal thiocarboxylate was generated using a 66 aa ThiS peptide, which functions as a sulfur donor for the biosynthesis of the thiazole moiety of thiamin. This was fused to the N terminus of the Sce VMA intein in the IMPACT-CN vector, and the resultant thioester was cleaved with ammonium sulfide. Five milligrams of ThiS-COSH peptide were purified per liter of culture.[94]



Scheme 9. Sequential condensation of protein fragments and the incorporation of a protein biosensor. Sequential condensation of synthetic peptides or biologically expressed polypeptides using a combination of intein-mediated expressed protein ligation and native peptide ligation allows incorporation of a Dns biosensor between the Abl kinase domains SH2 and SH3.[22] After condensation of the synthetic peptide with the C-terminal polypeptide, the fusion is treated with Factor Xa protease to generate a new N-terminal Cys. The fusion is then allowed to react with the N-terminal polypeptide to generate the mature protein product. The same technique can be used to introduce any unnatural amino acid or chemical reagent at any desired location in a protein. Xa: peptide with Factor Xa cleavage site.

4. Summary and Outlook

In less than ten years, protein splicing has shifted from being a biological oddity to a well-understood, widely observed process that has generated multiple protein-engineering applications. However, we are still unsure of the biological function of inteins. Some suggest that intein genes are selfish DNA elements that invade other genes with little penalty after splicing.[13, 35, 37, 48, 95] It has also been suggested that inteins control expression of the genes in which they reside.[11, 13, 48, 84, 95] However, the self-processing protein-splicing mechanism seems to leave little room for control, especially since splicing is rapid under physiological conditions. Natural precursors have not been detected in the native host, and no growth conditions have been found that modulate this process in the native organism. It has also been suggested that inteins may have played an important role in early evolution.[95] Trans-splicing and intein-mediated protein ligation suggest that inteins could have provided a means of trial-and-error domain shuffling to generate improved enzymes at the protein level, especially prior to the development of sophisticated DNA recombination pathways. Once DNA recombination becomes facile, intein genes could provide islands of homology for recombination between genes encoding smaller protein modules.

Although a general mechanism of protein splicing has been elucidated (Scheme 4), exceptions to this mechanism exist in nature. The largest void in our understanding of protein splicing is how the intein facilitates the four nucleophilic displacements. Creative approaches to determining the structure of precursors and intermediates are needed to solve this problem. For example, how do inteins enhance attack of the backbone carbonyl group by the Asn side-chain nitrogen atom, rather than the more common attack of the Asn side-

chain carbonyl group by the main-chain nitrogen atom? Since well-known chemical reactions make up the protein-splicing pathway, the development of applications utilizing inteins has been rapid. Although intein purification vectors were the first intein-derived application, intein-mediated expressed protein ligation and its related applications may prove to be the most important and widely used intein-derived techniques. The ability of intein vectors to chemoselectively generate large polypeptides of any sequence with α-thioesters or N-terminal cysteines circumvents the limits of solid-phase peptide synthesis. Inteins also provide a new target for drug discovery for several diseases. The last year has seen the development of many new protein-engineering applications which utilize IPL to synthesize cytotoxic proteins, add fluorescent tags to proteins, segmentally label proteins for NMR analysis, cyclize proteins, generate C-terminal thiocarboxylates, or introduce fluorophores into proteins. The list of intein-related applications will continue to expand as more chemists become aware of intein-derived tools.

We dedicate this review to the memory of Prof. Holger W. Jannasch (1927–1998) of the Woods Hole Oceanographic Institute, whose exploration of extreme environments led to the characterization of many microbes, including Thermococcus litoralis and Pyrococcus sp. GB-D, from which were cloned intein-containing genes that allowed us to define the protein-splicing mechanism. We thank our colleagues at New England BioLabs, Virginia Cornish and Tom Muir for valuable discussions, Karen Sandman, Karen Noren, Tom Evans, and Maurice Southworth for reading the manuscript, and Tom Muir for preparing Scheme 9.

Received: March 10, 1999 [A 331]

[1] F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner, M. Belfort, *Nucleic Acids Res.* **1994**, *22*, 1125–1127.

[2] F. B. Perler, M. Q. Xu, H. Paulus, *Curr. Opin. Chem. Biol.* **1997**, *1*, 292–299, and references therein.

[3] F. B. Perler, *Nat. Struct. Biol.* **1998**, *5*, 249–52, and references therein.

[4] M. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra, F. B. Perler, *Cell* **1993**, *75*, 1371–1377.

[5] M. Kawasaki, Y. Satow, Y. Anraku, *FEBS Lett.* **1997**, *412*, 518–520.

[6] S. Chong, K. S. Williams, C. Wotkowicz, M. Q. Xu, *J. Biol. Chem.* **1998**, *273*, 10567–10577.

[7] S. Chong, G. E. Montello, A. Zhang, E. J. Cantor, W. Liao, M. Q. Xu, J. Benner, *Nucleic Acids Res.* **1998**, *26*, 5109–5115.

[8] A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, Jr., F. B. Perler, *J. Bacteriol.* **1997**, *179*, 6378–6382.

[9] A. A. Cooper, Y. Chen, M. A. Lindorfer, T. H. Stevens, *EMBO J.* **1993**, *12*, 2575–2583.

[10] E. O. Davis, P. J. Jenner, P. C. Brooks, M. J. Colston, S. G. Sedgwick, *Cell* **1992**, *71*, 201–210.

[11] H. Paulus, *Chem. Soc. Rev.* **1998**, *27*, 375–386.

[12] M. W. Southworth, E. Adam, D. Panne, R. Byer, R. Kautz, F. B. Perler, *EMBO J.* **1998**, *17*, 918–926.

[13] V. Derbyshire, D. W. Wood, W. Wu, J. T. Dansereau, J. Z. Dalgaard, M. Belfort, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11466–11471.

[14] M. W. Southworth, K. Amaya, T. C. Evans, M. Xu, F. B. Perler, *BioTechniques* **1999**, *27*, 110–120.

[15] T. C. Evans, J. Benner, M. Q. Xu, *J. Biol. Chem.* **1999**, *274*, 3923–3926.

[16] S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus, M. Q. Xu, *Gene* **1997**, *192*, 271–281.

[17] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705–6710.

[18] T. C. Evans, J. Benner, M. Q. Xu, *Protein Sci.* **1998**, *7*, 2256–2264.

[19] K. Severinov, T. W. Muir, *J. Biol. Chem.* **1998**, *273*, 16205–16209.

[20] R. Xu, B. Ayers, D. Dowburn, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 388–393

[21] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, H. Nakamura, *J. Am. Chem. Soc.* **1998**, *120*, 5591–5592.

[22] G. J. Cotton, B. Ayers, R. Xu, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 1100–1101.

[23] P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebl, T. H. Stevens, *Science* **1990**, *250*, 651–657.

[24] R. Hirata, Y. Ohsumi, A. Nakano, H. Kawasaki, K. Suzuki, Y. Anraku, *J. Biol. Chem.* **1990**, *265*, 6726–6733.

[25] E. O. Davis, S. G. Sedgwick, M. J. Colston, *J. Bacteriol.* **1991**, *173*, 5653–5662.

[26] F. B. Perler, D. G. Comb, W. E. Jack, L. S. Moran, B. Qiang, R. B. Kucera, J. Benner, B. E. Slatko, D. O. Nwankwo, S. K. Hempstead, C. K. S. Carlow, H. Jannasch, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5577–5581.

[27] H. H. Gu, J. Xu, M. Gallagher, G. E. Dean, *J. Biol. Chem.* **1993**, *268*, 7372–7381.

[28] R. A. Hodges, F. B. Perler, C. J. Noren, W. E. Jack, *Nucleic Acids Res.* **1992**, *20*, 6153–6157.

[29] F. B. Perler, *Nucleic Acids Res.* **1999**, *27*, 346–347.

[30] F. B. Perler, G. J. Olsen, E. Adam, *Nucleic Acids Res.* **1997**, *25*, 1087–1093.

[31] S. Pietrokovski, *Protein Sci.* **1998**, *7*, 64–71.

[32] J. Z. Dalgaard, M. J. Moser, R. Hughey, I. S. Mian, *J. Comput. Biol.* **1997**, *4*, 193–214.

[33] X. Q. Liu, Z. Hu, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7851–7816.

[34] H. Fsihi, V. Vincent, S. T. Cole, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 3410–3415.

[35] M. Belfort, P. S. Perlman, *J. Biol. Chem.* **1995**, *270*, 30237–30240.

[36] A. E. Gorbalenya, *Nucleic Acids Res.* **1998**, *26*, 1741–1748.

[37] F. S. Gimble, J. Thorner, *Nature* **1992**, *357*, 301–306.

[38] M. Belfort, R. J. Roberts, *Nucleic Acids Res.* **1997**, *25*, 3379–3388, and references therein.

[39] J. Z. Dalgaard, A. J. Klar, M. J. Moser, W. R. Holley, A. Chatterjee, I. S. Mian, *Nucleic Acids Res.* **1997**, *25*, 4626–4638.

[40] S. Pietrokovski, *Protein Sci.* **1994**, *3*, 2340–2350.

[41] M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, *Nucleic Acids Res.* **1998**, *26*, 4409–4412.

[42] F. B. Perler, *Cell* **1998**, *92*, 1–4, and references therein.

[43] W. Grindl, W. Wende, V. Pingoud, A. Pingoud, *Nucleic Acids Res.* **1998**, *26*, 1857–1862.

[44] Z. He, M. Crist, H. Yen, X. Duan, F. A. Quiocho, F. S. Gimble, *J. Biol. Chem.* **1998**, *273*, 4607–4615.

[45] T. M. Hall, J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy, D. J. Leahy, *Cell* **1997**, *91*, 85–97.

[46] J. J. Lee, S. C. Ekker, D. P. von Kessler, J. A. Porter, B. I. Sun, P. A. Beachy, *Science* **1994**, *266*, 1528–1537.

[47] E. V. Koonin, *Trends Biochem. Sci.* **1995**, *20*, 141–142.

[48] a) J. A. Porter, S. C. Ekker, W. J. Park, D. P. von Kessler, K. E. Young, C. H. Chen, Y. Ma, A. S. Woods, R. J. Cotter, E. V. Koonin, P. A. Beachy, *Cell* **1996**, *86*, 21–34; b) P. A. Beachy, M. K. Cooper, K. E. Young, D. P. von Kessler, W. Park, T. M. T. Hall, D. J. Leahy, J. A. Porter, *Cold Spring Harbor Symp. Quant. Biol.* **1997**, *62*, 191–204.

[49] T. R. Burglin, *Curr. Biol.* **1996**, *6*, 1047–1050, and references therein.

[50] A. R. Fersht, *Enzyme Structure and Mechanism*, W. H. Freeman, Reading, **1977**.

[51] C. F. Barbas III, J. R. Matos, J. B. West, C. H. Wong, *J. Am. Chem. Soc.* **1988**, *110*, 5162–5166.

[52] T. Nakatsuka, T. Sasaki, E. T. Kaiser, *J. Am. Chem. Soc.* **1987**, *109*, 3808–3810.

[53] L. Abrahmsen, J. Tom, J. Burnier, K. A. Butcher, A. Kossiakoff, J. A. Wells, *Biochemistry* **1991**, *30*, 4151–4159.

[54] A. Warshel, G. Naray-Szabo, F. Sussman, J. K. Hwang, *Biochemistry* **1989**, *28*, 3629–3637.

[55] D. R. Corey, C. S. Craik, *J. Am. Chem. Soc.* **1992**, *114*, 1784–1790.

[56] K. Iwai, T. Ando, *Methods Enzymol.* **1967**, *11*, 262–282.

[57] T. C. Bruice, S. J. Benkovic in *Bioorganic Mechanisms*, *Vol. 1* (Eds.: R. Breslow, M. Karplus), W. A. Benjamin, New York, **1966**, pp. 266 – 297.
[58] W. P. Jencks, S. Cordes, J. Carriulo, *J. Biol. Chem.* **1960**, *235*, 3608 – 3614.
[59] A. A. Cooper, T. H. Stevens, *BioEssays* **1993**, *15*, 667 – 673.
[60] M. Xu, F. B. Perler, *EMBO J.* **1996**, *15*, 5146 – 5153.
[61] N. D. Clarke, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11 084 – 11 088.
[62] C. J. Wallace, *Protein Sci.* **1993**, *2*, 697 – 705.
[63] M. Xu, D. G. Comb, H. Paulus, C. J. Noren, Y. Shao, F. B. Perler, *EMBO J.* **1994**, *13*, 5517 – 5522.
[64] Y. Shao, H. Paulus, *J. Pept. Res.* **1997**, *50*, 193 – 198.
[65] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776 – 779.
[66] S. Chong, Y. Shao, H. Paulus, J. Benner, F. B. Perler, M. Q. Xu, *J. Biol. Chem.* **1996**, *271*, 22 159 – 22 168.
[67] Y. Shao, M. Q. Xu, H. Paulus, *Biochemistry* **1995**, *34*, 10 844 – 10 850.
[68] T. V. Brennan, S. Clarke, *Int. J. Pept. Protein Res.* **1995**, *45*, 547 – 553, and references therein.
[69] C. E. Voorter, W. A. de Haard-Koekman, P. J. van den Oetelaar, H. Bloemendal, W. W. de Jong, *J. Biol. Chem.* **1988**, *263*, 19 020 – 19 023.
[70] A. V. Klotz, *Bioorg. Chem.* **1993**, *21*, 83 – 94.
[71] T. Geiger, S. Clarke, *J. Biol. Chem.* **1987**, *262*, 785 – 794.
[72] S. Clarke, R. C. Stephenson, J. D. Lowenson in *Stability of Protein Pharmaceuticals, Part A* (Eds.: T. J. Ahern, M. C. Manning), Plenum, New York, **1992**, pp. 1 – 29, and references therein.
[73] R. C. Stephenson, S. Clarke, *J. Biol. Chem.* **1989**, *264*, 6164 – 6170.
[74] R. Hirata, Y. Anraku, *Biochem. Biophys. Res. Commun.* **1992**, *188*, 40 – 47.
[75] P. Vanderslice, W. C. Copeland, J. D. Robertus, *J. Biol. Chem.* **1988**, *263*, 10 583 – 10 586.
[76] S. Nogami, Y. Satow, Y. Ohya, Y. Anraku, *Genetics* **1997**, *147*, 73 – 85.
[77] M. Kawasaki, S. Nogami, Y. Satow, Y. Ohya, Y. Anraku, *J. Biol. Chem.* **1997**, *272*, 15 668 – 15 674.
[78] J. N. Higaki, L. B. Evnin, C. S. Craik, *Biochemistry* **1989**, *28*, 9256 – 9263.
[79] C. Guan, T. Cui, V. Rao, W. Liao, J. Benner, C. L. Lin, D. Comb, *J. Biol. Chem.* **1996**, *271*, 1732 – 1737.
[80] Y. Shao, M. Q. Xu, H. Paulus, *Biochemistry* **1996**, *35*, 3810 – 3815.
[81] M. W. Southworth, F. B. Perler, unpublished results.
[82] J. A. Porter, D. P. von Kessler, S. C. Ekker, K. E. Young, J. J. Lee, K. Moses, P. A. Beachy, *Nature* **1995**, *374*, 363 – 366.
[83] C. Huang, S. Wang, L. Chen, C. Lemieux, C. Otis, M. Turmel, X. Q. Liu, *Mol. Gen. Genet.* **1994**, *244*, 151 – 159.
[84] H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226 – 9231.
[85] L. Chen, F. B. Perler, unpublished results.
[86] X. Duan, F. S. Gimble, F. A. Quiocho, *Cell* **1997**, *89*, 555 – 564.
[87] T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs, Jr., J. C. Sacchettini, *Nat. Struct. Biol.* **1998**, *5*, 31 – 36.
[88] B. L. Stoddard, S. Pietrokovski, *Nat. Struct. Biol.* **1998**, *5*, 3 – 5.
[89] S. N. Cook, W. E. Jack, X. Xiong, L. E. Danley, J. A. Ellman, P. G. Schultz, C. J. Noren, *Angew. Chem.* **1995**, *107*, 1736 – 1737; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1629 – 1630.
[90] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182 – 188.
[91] K. V. Mills, B. M. Lew, S. Jiang, H. Paulus, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3543 – 3548.
[92] B. M. Lew, K. V. Mills, H. Paulus, *J. Biol. Chem.* **1998**, *273*, 15 887 – 15 890.
[93] a) T. C. Evans, Jr., J. Benner, M. Q. Xu, *J. Biol. Chem.* **1999**, *274*, 18 359 – 18 363; b) J. A. Camarero, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 5597 – 5598.
[94] C. Kinsland, S. V. Taylor, N. L. Kelleher, F. W. McLafferty, T. P. Begley, *Protein Sci.* **1998**, *7*, 1839 – 1842.
[95] F. B. Perler, *Trends Biochem. Sci.* **1999**, *24*, 209 – 210, and references therein.