Bioorganic Chemistry **29,** 119–129 (2001)
doi:10.1006/bioo.2001.1203, available online at http://www.idealibrary.com on IDEAL®

MINIREVIEW

Inteins as Enzymes

Henry Paulus[1]

Boston Biomedical Research Institute, 64 Grove Street, Watertown, Massachusetts 02472 and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Received September 20, 2000

The term "self-catalyzed" as applied to protein processing reactions might be considered a contradiction, since catalysis implies that the catalyst is regenerated without change. However, as our understanding of protein autoprocessing reactions such as protein splicing advances, it is becoming clear that they have many of the hallmarks of enzymatic reactions. In this review, we will examine the properties of protein splicing elements, or inteins, and show how these can be understood in terms of enzyme catalysis, both with respect to substrate specificity and the stabilization of reactive intermediates.

Key Words: protein splicing; hedgehog proteins; glycosylasparaginases; pyruvoyl enzymes.

INTRODUCTION

In recent years, a number of protein processing reactions have been described that are distinguished by the fact that they occur in the absence of accessory proteins and involve as their first step the N-S or N-O acyl rearrangement of a peptide bond involving a cysteine, serine, or threonine residue (Fig. 1). Since such peptide bond rearrangements do not occur spontaneously at neutral pH, these protein processing reactions have been described as "self-catalyzed" in spite of the fact that the "catalyst" is not regenerated in a form capable of promoting additional reaction cycles. In the case of protein splicing, the classical experiments by Xu and coworkers (*1*) established that all information required for catalysis resides in the intein, the intervening sequence that is excised in this process. Inteins, which range in size from 146 to 609 amino acids (*2*) and are inserted into host proteins before a cysteine, serine, or threonine residue, catalyze the first three of the reactions shown in Fig. 2 (for a recent review, see Ref. *3*). In this article, we will look at protein splicing from the point of view

[1] Address correspondence and reprint requests to author at Boston Biomedical Research Institute, 64 Grove Street, Watertown, MA 02472-2829. Fax: (617) 972-1753. E-mail: paulus@bbri.org.

0045-2068/01 $35.00

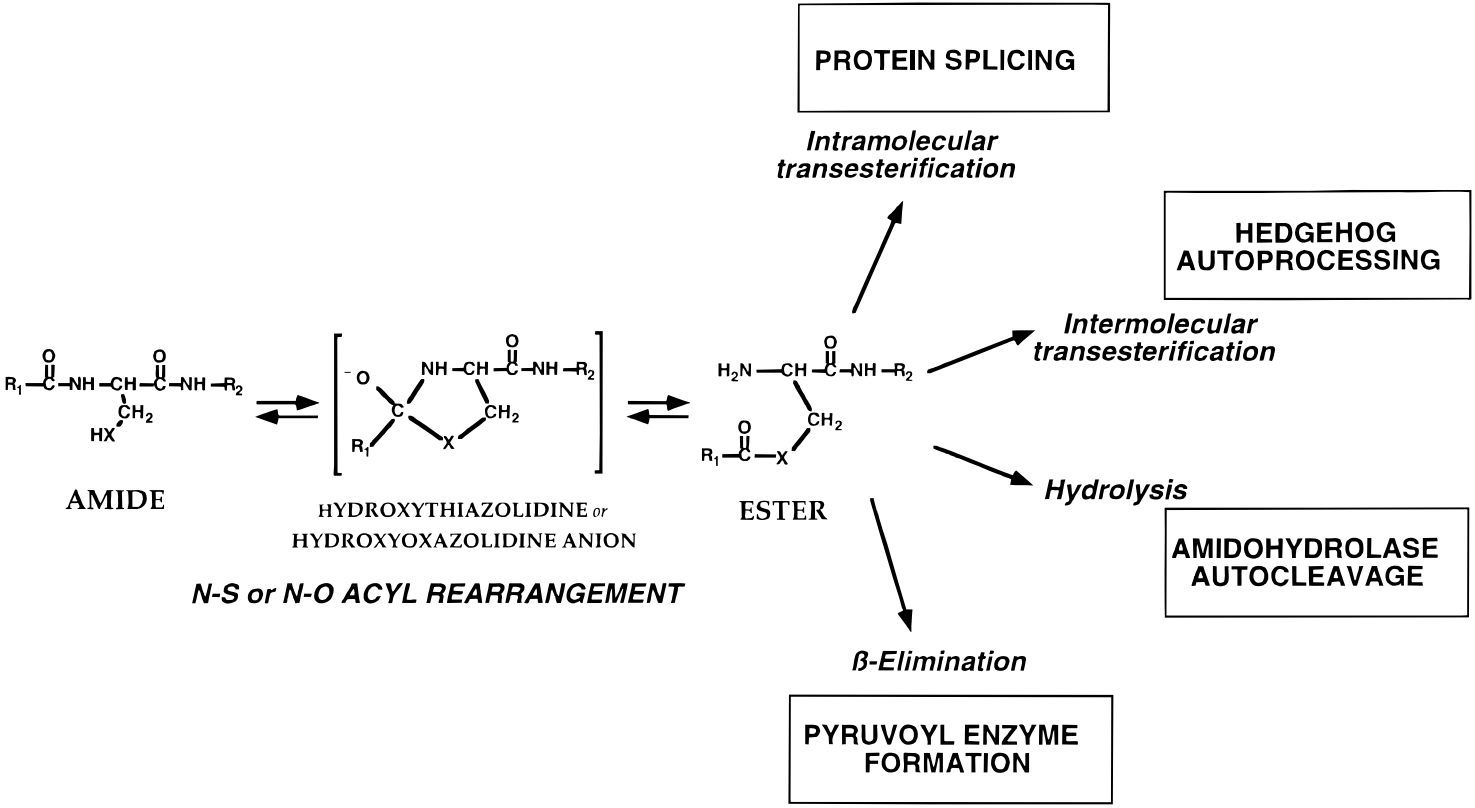


FIG. 1. Protein autoprocessing reactions initiated by self-catalyzed N-O or N-S acyl rearrangements.

of classical enzymology in order to gain a better understanding of its mechanism and specificity.

PROTEIN SPLICING VIEWED AS AN ENZYMATIC REACTION

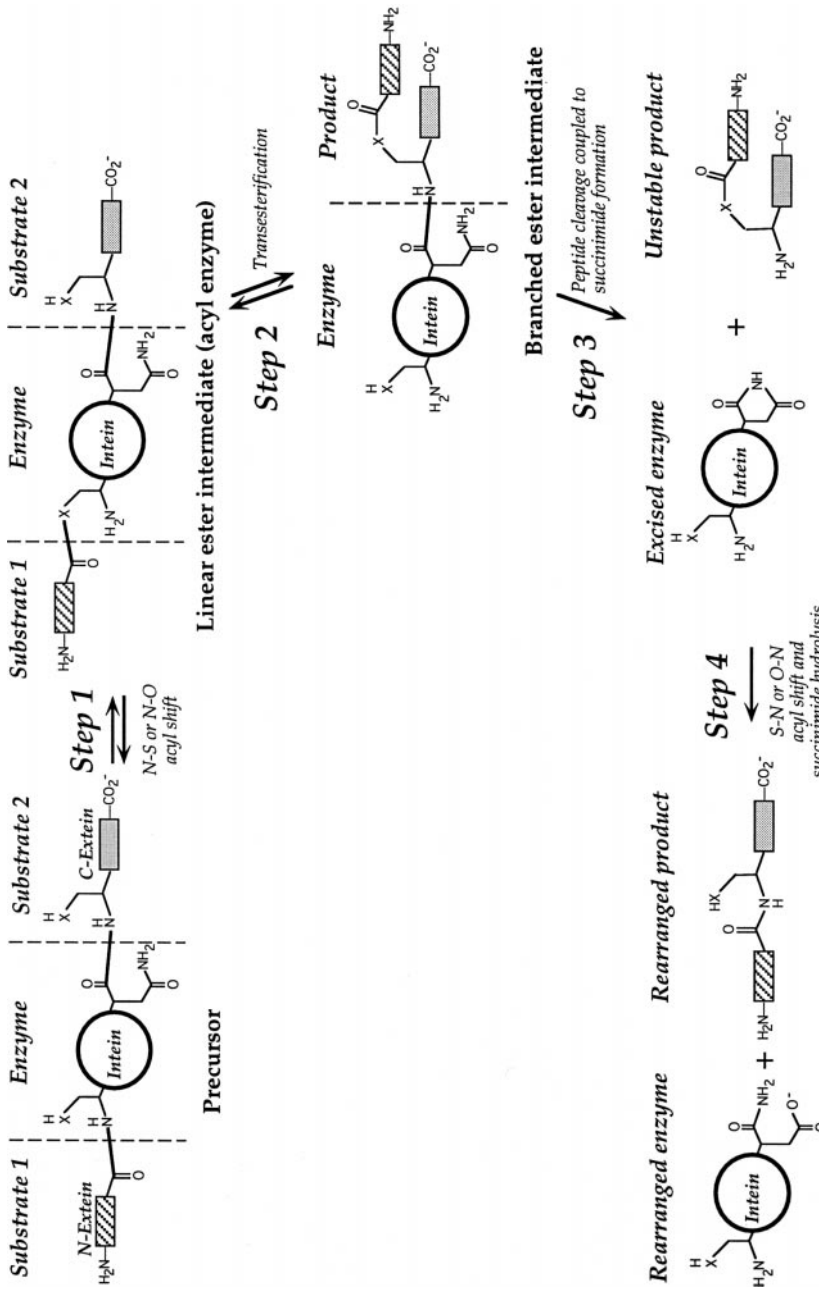
Let us begin by examining the reactions catalyzed by inteins from the point of view of the enzymologist. Protein splicing involves the action of an enzyme—the intein—on two substrates—the N-extein and the C-extein—which leads to the breaking of the peptide bonds linking the intein and the exteins and the formation of a new ester bond between the exteins, as indicated in Fig. 2. The ester bond subsequently rearranges spontaneously to a peptide bond. In step 1, the nucleophilic side chain of the N-terminal residue of the enzyme attacks the C-terminal carbonyl carbon of substrate 1, forming an acyl-enzyme intermediate (the linear ester intermediate) with the concomitant scission of the peptide bond between the enzyme and substrate 1. In step 2, the enzyme catalyzes the transfer of substrate 1 to the nucleophilic side chain of the N-terminal residue of substrate 2, yielding an ester product that is still linked to the C-terminus of the enzyme by a peptide bond (the branched ester intermediate).[2] In step 3, the peptide bond linking the product to the enzyme is broken by the nucleophilic attack of the β-amido group of the C-terminal asparagine of the enzyme on its α-carbonyl carbon. The product is released as a thermodynamically and kinetically unstable peptide ester (*5*), which rearranges spontaneously to the corresponding amide (step 4). The first two steps in the protein splicing pathway are the equivalent of a double-displacement mechanism, with the complication that the enzyme is initially joined to the substrates by covalent bonds, one of which is broken in the course of the reactions. The third step is a mechanistically unrelated nucleophilic displacement reaction to yield an intein with a C-terminal aminosuccinimide residue. It has not yet been determined whether the hydrolysis of the aminosuccinimide moiety in step 4 is spontaneous or catalyzed by the intein.[3]

SUBSTRATE SPECIFICITY IN PROTEIN SPLICING

The fact that inteins are often found in closely related proteins, presumably owing to horizontal transfer mediated by the homing endonucleases that are embedded in most inteins (*7*,*8*), provides an opportunity to examine substrate specificity in protein splicing. The amino acid sequences of inteins inserted at allelic sites in a particular host protein have usually diverged much more than those of the host proteins themselves. For example, the DNA gyrase A subunits of six different mycobacterial species have an average sequence identity of 96%, but the inteins that are inserted at equivalent sites in these proteins are on average only 65% identical. This divergence implies evolution of the inteins subsequent to their insertion to optimize protein splicing in

[2] In some inteins, formation of the acyl-enzyme intermediate (step 1) is circumvented by the direct transfer of substrate 1 to the nucleophilic side chain of the N-terminal residue of substrate 2, thus combining reactions 1 and 2 into a single step (4). In the terms of our discussion, this alternative mechanism does not involve an acyl-enzyme intermediate and thus represents a single-displacement reaction mechanism, contrasting with the double-displacement mechanism illustrated in Fig. 2.

[3] This question could be answered by analysis of the excised intein. Spontaneous hydrolysis would lead to an equilibrium mixture of C-terminal asparagine and isoasparagine residues, whereas intein-catalyzed hydrolysis would produce only one of these isomers (6).


Substrate 1
Enzyme
Substrate 2
N-Extein
Intein
C-Extein
Step 1
N-S or N-O acyl shift
Precursor
Linear ester intermediate (acyl enzyme)
Step 2
Transesterification
Enzyme
Product
Intein
Branched ester intermediate
Step 3
Peptide cleavage coupled to succinimide formation
Excised enzyme
Unstable product
Intein
Rearranged enzyme
Rearranged product
Step 4
S-N or O-N acyl shift and succinimide hydrolysis
Intein

the context of the host protein. One would expect this optimization to involve specific interactions of the intein with the host proteins that improve the efficiency of protein splicing under physiological conditions and to minimize the occurrence of side reactions. Evidence for such specificity was obtained by a detailed study of the GyrA intein of *Mycobacterium xenopii* (*9*). When a fusion protein containing the GyrA intein together with 64–65 GyrA residues at its N- and C-termini is expressed in *Escherichia coli*, protein splicing occurs with high efficiency *in vivo* (86% at 37°C and 100% at 16°C). When the native GyrA residues at the intein C-terminus are replaced by a foreign polypeptide, protein splicing becomes temperature-sensitive (only 2% protein splicing at 37°C but 100% splicing at 16°C). When the native GyrA residues at the intein N-terminus are replaced by a foreign sequence, no protein splicing is seen either at 37 or 16°C. These observations suggest that productive protein splicing requires interactions between the intein and adjacent extein residues. Experiments with the split DnaE intein of *Synechocystis* sp. PCC 6803 also showed that efficient protein splicing requires flanking extein sequences, with optimal splicing activity observed with five DnaE residues adjacent to the intein N-terminus and three DnaE residues at the C-terminus (*10*).

An interesting aspect of these specificity studies is that the various reactions in the protein splicing pathway are differentially affected by the presence of adjacent extein residues. Although the absence of native GyrA residues at the N-terminus of the *M. xenopii* GyrA intein completely prevents protein splicing, reactions 1 and 3 of the protein splicing pathway can still occur, leading—under appropriate conditions—to cleavage of either of the scissile bonds linking the intein to the exteins (*9*). Indeed, protein splicing mediated by inteins inserted between foreign exteins is generally less efficient than splicing in the native context and is accompanied by considerable cleavage at the N- and C-terminal splice junctions (e.g., *1,11,12*). This is reminiscent of the uncoupling of the steps in complex enzymatic reaction pathways that sometimes occurs in the presence of substrate analogs. For example, in the presence of 5-oxo-L-proline analogs such as L-2-imidazolidone-4-carboxylate, 5-oxo-L-prolinase catalyses ATP hydrolysis without amino acid cleavage (*13*). Similarly, when presented with noncognate amino acids, some aminoacyl-tRNA synthetases catalyze the hydrolysis of the aminoacyl adenylate intermediates rather than the transfer of the amino acid to tRNA (*14*). In the case of protein splicing, it appears that the transesterification reaction (step 2) is most sensitive to perturbation of substrate structure, with the result that the cleavage reactions are uncoupled from transesterification and occur autonomously. One might expect step 2 of protein splicing, in which the N-extein is transferred to the nucleophilic side chain of the N-terminal residue of the C-extein, to involve relatively extensive intein–extein (enzyme–substrate) interactions and thus

FIG. 2. An enzymologist's view of the mechanism of protein splicing. The unspliced precursor and the splicing intermediates are partitioned by the dotted lines into *enzyme* and *substrate* moieties as discussed in the text. The amino acid residues that participate directly in the chemical transformations are shown (X = O or S) and the remainder of the intein and exteins are shown by circles or boxes, which are not to scale. The scissile peptide bonds linking the intein to the exteins are shown by the elongated heavy line.