Putting protein splicing to work

Frederick S Gimble

Several protein processing events that involve related chemical mechanisms have been observed in nature. Now, new methods have been developed, based on the same chemical reactions, that permit proteins to be modified in ways that were not previously possible.

Address: Center for Macromolecular Design, Institute of Biosciences and Technology and Department of Biochemistry and Biophysics, 2121 West Holcombe Boulevard, Texas A&M University, Houston, Texas 77030, USA.

E-mail: fgimble@ibt.tamu.edu

Chemistry & Biology October 1998, 5:R251–R256
http://biomednet.com/elecref/10745521005R0251

Introduction

The unexpected discovery of protein splicing in 1990 [1,2] demonstrated that the transfer of information from gene to protein can be profoundly influenced by post-translational modification. Protein splicing is a multi-step process whereby an internal polypeptide segment, termed the intein, of a protein precursor autocatalytically breaks two peptide bonds to excise itself and uses a peptide bond to join the amino- and carboxy-terminal precursor segments, the amino and carboxyl exteins, to form a second protein (Figure 1a). Intein sequences have been identified in archaea, eubacteria and eukarya. Several inteins occur as bifunctional proteins that not only catalyze protein splicing but also function as rare-cutting endonucleases that initiate the mobility of the intein gene, in a process called intein homing [3].

In the protein-splicing pathway (Figure 1a), the normally stable peptide bond at the amino-terminal splice junction is first converted to a more reactive ester or thioester linkage by acyl rearrangement [4]. This occurs by nucleophilic attack of the peptide bond carbonyl by an adjacent serine or cysteine sidechain. Next, a branched intermediate with two amino termini is generated when the newly formed ester bond is cleaved during a transesterification reaction by attack from a nucleophilic residue (cysteine, serine or threonine) located at the carboxy-terminal splice site. Release of the intein occurs when an asparagine sidechain at the intein carboxyl terminus attacks the carboxy-terminal scissile peptide bond. Finally, the ester linkage in the fused extein product rapidly rearranges to form the thermodynamically more stable peptide bond.

The first examples of protein splicing to be described all utilized intramolecular transesterification steps, but the recent identification of a split intein in the cyanobacterium *Synechocystis* reveals that *trans* splicing also occurs in nature (Figure 1b). Papers by Gorbalenya [5] and by Wu *et al.* [6] noted that two open reading frames, separated on the chromosome by 745 kilobases (kb), both contained similarity to parts of inteins and to parts of the *dnaE* gene, which encodes the DNA polymerase III α subunit. The two gene products of these open reading frames comprise a split intein in which one protein contains residues 1–744 of the DnaE protein fused to the first 123 amino acids of an intein, and the second contains 36 carboxy-terminal intein residues fused to 423 carboxy-terminal amino acids of DnaE. Furthermore, this split intein resembles other small contiguous inteins, termed mini-inteins, in that it lacks an endonuclease region. Remarkably, when the two gene fusions are expressed in *Escherichia coli*, protein *trans* splicing occurs *in vivo* to generate the ligated DnaE protein [6]. It is presumed that the two intein protein fragments (the amino and carboxyl inteins) fold together properly in the cell to yield protein-splicing activity.

A further variation on the protein-splicing theme, albeit one far less complex than intein splicing, is the processing of the *Drosophila* Hedgehog protein, a secreted protein responsible for patterning in embryonic segments and imaginal discs. Hedgehog is a two-domain protein, and the peptide bond connecting the domains is converted to a thioester linkage by an acyl rearrangement reaction analogous to that in intein splicing (Figure 1c) [7]. Indeed, the Hedgehog carboxy-terminal domain, which mediates the rearrangement, is structurally similar to a splicing domain from a yeast intein [8,9]. In contrast to most instances of intein splicing, however, intramolecular transesterification does not occur; instead, a cholesterol molecule acts as nucleophile and breaks the bond between the two domains, becoming covalently attached to the end of the amino-terminal signaling domain in the process [7]. Tethering of the hydrophobic cholesterol to the amino-terminal domain binds the protein tightly to the cell surface and effects signaling.

The original hope that practical applications could be developed from protein splicing has now been realized with the development of methods based on protein splicing chemistry that permit the joining of polypeptide segments *in vitro*, as well as the introduction of detection probes and specific modifications into proteins at defined locations. What is interesting is that these methods parallel the different variations of protein splicing observed in nature. One approach is analogous to *trans* splicing *in vivo* and makes use of split inteins to reconstruct proteins. The underlying basis of the second approach is more related to Hedgehog processing, and is a semi-synthesis method that

Figure 1

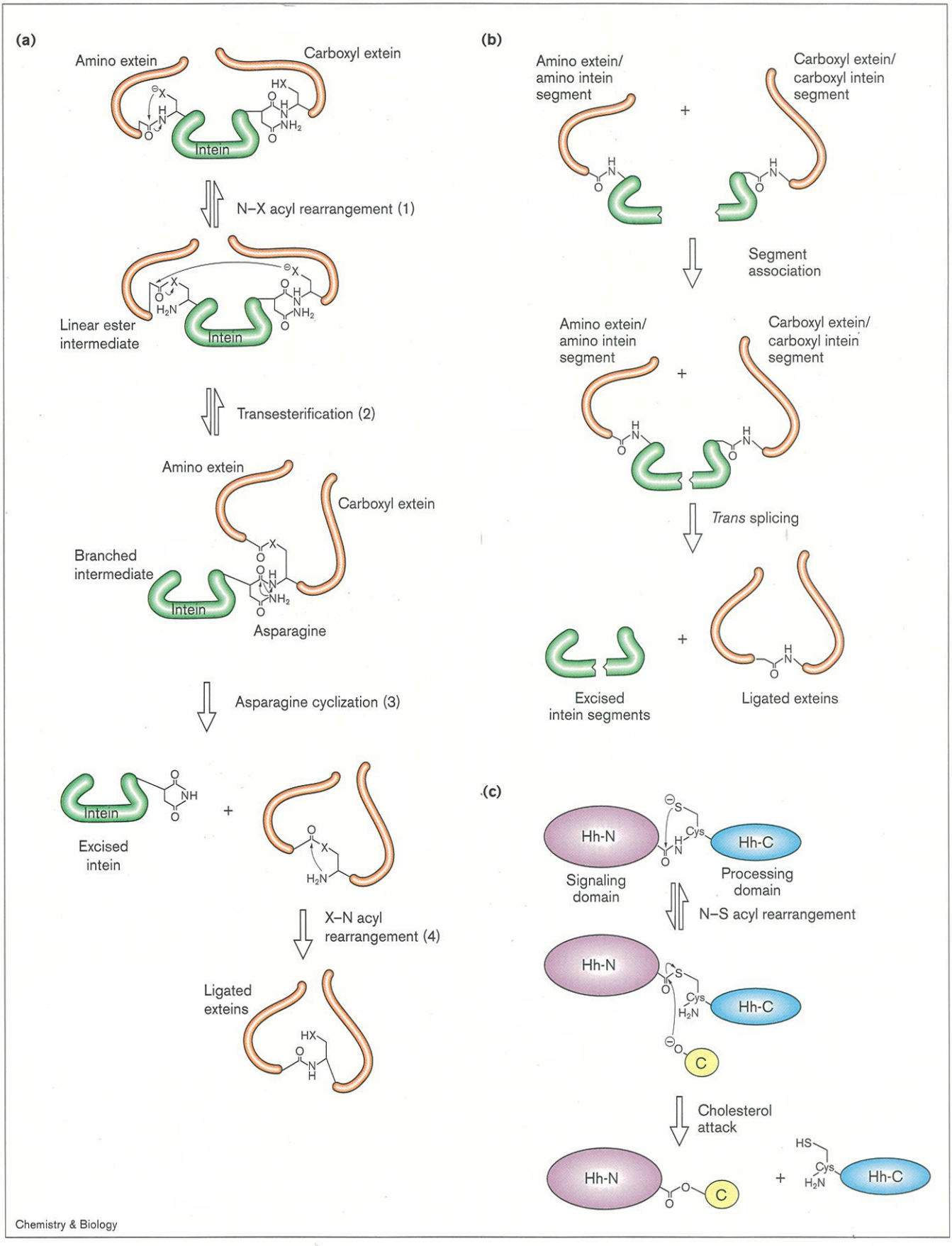

(a)
Amino extein
Carboxyl extein
Intein
N–X acyl rearrangement (1)
Linear ester intermediate
Transesterification (2)
Amino extein
Carboxyl extein
Branched intermediate
Asparagine
Asparagine cyclization (3)
Excised intein
X–N acyl rearrangement (4)
Ligated exteins
(b)
Amino extein/ amino intein segment
Carboxyl extein/ carboxyl intein segment
Segment association
Trans splicing
Excised intein segments
Ligated exteins
(c)
Hh-N
Hh-C
Signaling domain
Processing domain
N–S acyl rearrangement
Cholesterol attack
Chemistry & Biology

Figure 1

Protein processing pathways *in vivo*. **(a)** Protein splicing. Acyl rearrangement of the intein amino-terminal serine or cysteine converts the peptide bond to an ester or thioester. This ester linkage is attacked by a serine, threonine or cysteine sidechain located at the amino-terminal position of the carboxyl extein to yield a branched-ester intermediate. Release of the intein occurs when the asparagine at the intein carboxyl terminus cyclizes and forms a succinimide ring, which is subsequently hydrolyzed to yield asparagine or isoasparagine. Finally, a second, spontaneous X–N acyl rearrangement forms a new peptide bond between the two fused exteins. The oxygen or sulfur atom of the reactive serine, threonine or cysteine is represented by an 'X'. **(b)** *Trans* splicing. It is presumed that *trans* splicing requires that the amino- and carboxy-terminal segments of the split intein first fold together *in vivo*. In addition, folding of the two extein polypeptides is likely to take place. After the intein structure forms, protein splicing is believed to follow the same pathway as that of contiguous inteins. **(c)** Hedgehog processing. The peptide bond between the signaling and processing domains of the Hedgehog protein is replaced by a thioester linkage following an acyl rearrangement identical to the first step of the protein-splicing pathway. The newly formed thioester is subject to nucleophilic attack by the 3β-hydroxyl group of a cholesterol molecule. The end result is that the two domains are separated, and the cholesterol molecule is covalently attached to the carboxyl terminus of the signaling domain. The yellow molecule labeled 'C' is cholesterol.

permits the ligation of a chemically synthesized carboxy-terminal protein segment containing engineered modifications to a recombinant amino-terminal segment.

Ligation of polypeptide segments *in vivo* and *in vitro* by *trans* splicing

Scientists have long sought the ability to engineer natural and unnatural modifications at precise and unique locations in proteins. Use of a split intein in a *trans* splicing system has the potential to permit the selective labeling or modification (e.g., phosphorylation, glycosylation or acetylation) of one region only of a protein. *Trans* splicing *in vivo* was engineered in the laboratory initially by introducing translational termination and initiation codons in-frame into the middle of the normally contiguous *Mycobacterium tuberculosis recA* intein positioned between a maltose-binding protein (MBP) amino extein and a carboxyl extein polypeptide containing a His-tag (His_6) [10]. Splicing of the split intein was confirmed by detection of the expected 48 kDa MBP–His_6 splicing product. Interestingly, splicing is more efficient using a split mini-intein version lacking the endonuclease region, which might form a structure analogous to the *Synechocystis dnaE* split intein. A mini-intein engineered from the *Synechocystis dnaB* intein also *trans* splices *in vivo*, suggesting that many inteins can fold from discontinuous segments and still maintain splicing activity [11].

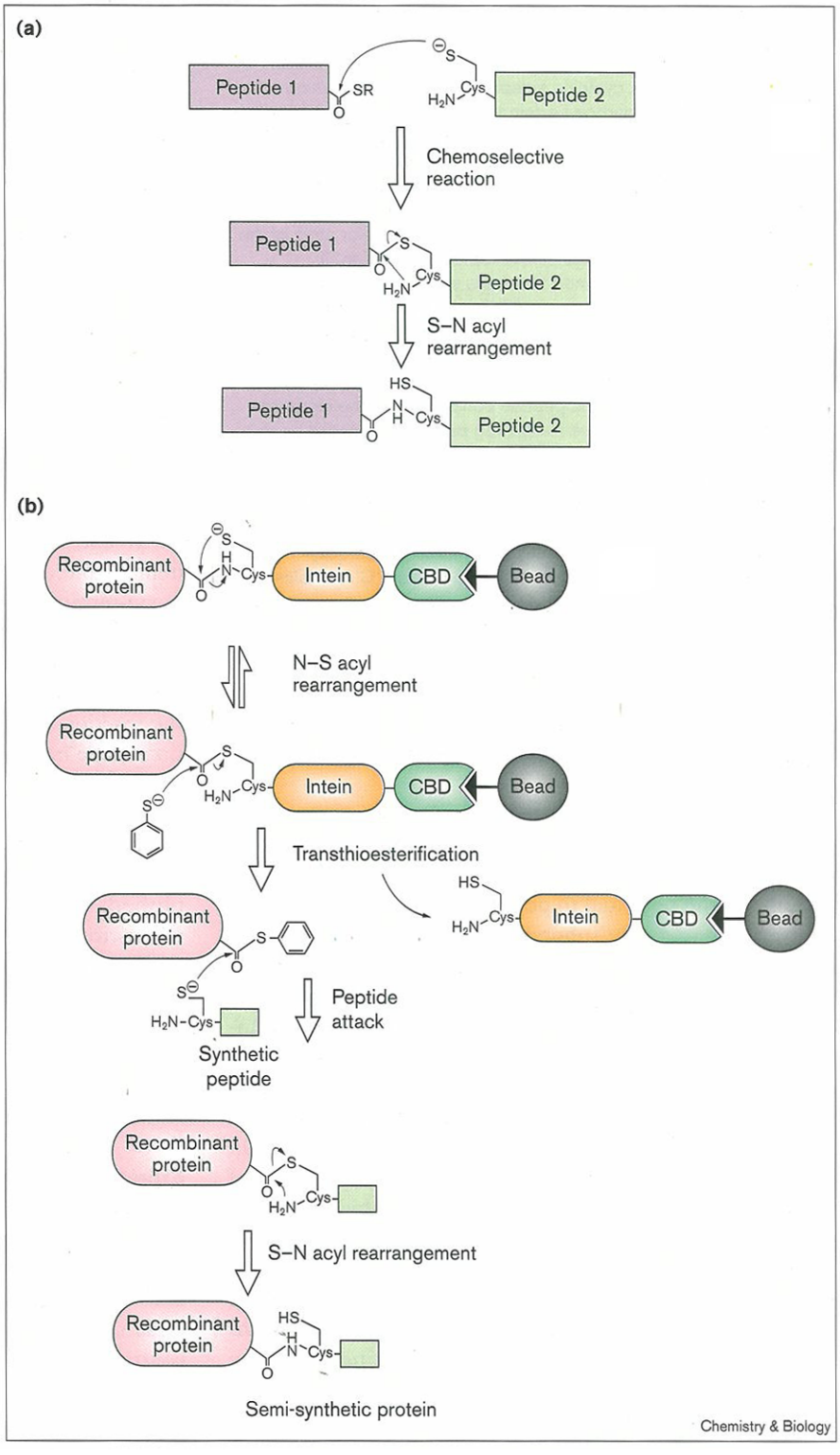
Two groups successfully reconstituted a *trans*-splicing system *in vitro*, which will make it easier to use chemically modified or synthetic exteins in the method. Mills *et al.* [12] extended their *trans*-splicing work *in vivo* by purifying the two segments of the split *M. tuberculosis* mini-intein. Southworth *et al.* [13] constructed similar split-intein polypeptides derived from a DNA polymerase intein in *Pyrococcus*, a thermophilic archaebacterium, but did not remove the endonuclease region from the intein. To reconstitute splicing activity, the two split-intein fragments first had to be combined and pre-incubated in urea, a protein denaturant, and then renatured by removal of the denaturant to form a splicing-competent complex. Presumably, reassociation of the amino and carboxyl inteins must occur from a partially or completely denatured state, and proper folding of an intein structure that has protein-splicing activity takes place during renaturation. Once reconstituted, the *Pyrococcus* intein is active in urea concentrations up to 6 M [13], but the *M. tuberculosis* split intein loses all splicing activity at urea concentrations above 2 M [12], perhaps reflecting a greater stability of the thermostable protein to denaturants. Depending on the conditions, splicing efficiency approached 80% or higher. Pre-incubation in denaturant is not required if the exteins are short, given that a 9 kDa amino extein will splice to a three-residue carboxyl extein [13]. Paulus and coworkers [14] have also demonstrated that replacement of the recombinant carboxyl-terminal intein–extein segment by a chemically synthesized peptide containing a 35-residue carboxyl intein and a short two-residue carboxyl extein still permits *trans* splicing to occur. The ability to use a synthetic peptide in the reaction will potentially allow a wide range of modifications to be introduced into a protein by *trans* splicing.

The generation of an active protein by *trans* splicing under native conditions remains to be achieved, but the studies reviewed here lead one to believe that this will be accomplished. The requirement for denaturation during the reconstitution step will probably require refolding of most spliced proteins to yield an active conformation, and this will not be possible for all proteins. To favor a successful outcome for any particular protein, several parameters of the reaction will have to be optimized, including choice of intein, positioning of the intein within the target protein and concentration of denaturant during reconstitution.

Expressed protein ligation

Current techniques for introducing unnatural molecules at specific locations have been limited in their use for a variety of reasons. For example, insertion of unnatural amino acids into proteins *in vitro* using chemically misacylated tRNAs results in low yields of modified protein. By contrast the total chemical synthesis of a protein allows one to introduce unnatural amino acids freely but is limited to short polypeptides. Somewhat larger polypeptides can be constructed using chemical ligation methods, however. The so-called 'native chemical ligation' technique permits two synthetic fragments to be joined by a peptide bond by means of a precise reaction between two complementary reactive groups [15]. It is interesting to note that the same chemistry is used as in steps 2 and 4 of the protein-splicing pathway (compare Figure 2a with Figure 1a). One peptide

Figure 2



Synthetic and semi-synthetic ligation methods. **(a)** Native chemical ligation. A synthetic peptide containing a thioester at the α-carboxyl group undergoes nucleophilic attack by the sidechain of a cysteine residue at the amino terminus of the second peptide. The thioester intermediate spontaneously undergoes an S–N acyl rearrangement to yield a peptide bond at the ligation site. **(b)** Expressed protein ligation. The protein of interest is expressed in bacteria as an in-frame fusion with a modified intein and a chitin-binding domain (CBD) and is affinity purified from contaminating proteins on chitin resin. As in protein splicing, spontaneous acyl rearrangement generates a thioester at the junction of the target protein and the intein. Addition of a small thiol compound (thiophenol is shown here) cleaves the target protein from the intein–CBD fusion, which remains attached to the resin, and generates a new thioester at the carboxyl terminus of the target protein. This new thioester-tagged intermediate can be eluted from the column in purified form and reacted with a synthetic peptide containing a cysteine at its amino-terminal position to generate the ligation product [18]. Alternatively, cleavage from the intein fusion precursor and ligation can be performed simultaneously *in situ* [16,17]. Finally, a spontaneous S–N acyl rearrangement converts the thioester linkage between the target protein and the peptide into a peptide bond.

fragment containing a pre-formed thioester at its carboxyl terminus is coupled by transthioesterification to a second peptide containing a cysteine residue at its amino terminus, ultimately forming a peptide bond between the two peptides. Finally, novel semisynthetic methods have been used to couple synthetic peptides to polypeptide fragments derived from natural proteins to generate larger proteins. But the drawback with these approaches has been in

obtaining the appropriate protein-cleavage fragments to use for ligation.

Two groups [16–18] have devised a clever strategy called 'expressed protein ligation' that takes advantage of the protein-splicing machinery to solve some of the problems associated with the existing synthetic methods (Figure 2b). Expressed protein ligation permits one to ligate a synthetic peptide to the carboxyl terminus of a recombinant protein. The technique is an extension of native chemical ligation, and its major benefit is that it can introduce the required carboxy-terminal thioester at a specific location within a recombinant protein of any size. Both groups express the target protein that will be ligated to the synthetic peptide as an in-frame fusion to an intein followed by a chitin-binding domain (CBD), which allows affinity purification of the fusion protein on a chitin column (Figure 2b). The vector constructs that facilitate making these fusion proteins were originally developed to permit the rapid purification of any target protein [19] and are commercially available. The intein used is a mutant version that is able to undergo the acyl rearrangement at the amino-terminal splice junction between the target protein and the intein, but cannot complete the splicing reaction due to a mutation at the carboxy-terminal junction. As a consequence, an intermediate forms containing a thioester linkage between the target protein and the mutant intein. Addition of a thiol-containing molecule, such as thiophenol [16,17] or 2-mercaptoethanesulfonic acid (MESNA) [18], induces intein-mediated cleavage from the column through a trans-thioesterification reaction. The newly formed thioester-tagged protein is then available to react with a synthetic peptide containing an amino-terminal cysteine residue. Replacement of the intein with the smaller thiol-containing molecule at the carboxy-terminal tail of the target protein is necessary because the peptide fails to react directly with the intein-linked thioester, presumably because of steric hindrance. Xu and coworkers [18] use an intein derived from the *M. xenopi* GyrA protein that is efficiently cleaved by MESNA, allowing the virtually complete recovery of the protein as the thioester-tagged intermediate prior to ligation. The yeast intein used by Muir's group [16,17] is cleaved less well, but efficient ligation is achieved nonetheless (> 90%) by performing the thiol-induced cleavage and ligation reactions simultaneously on the column.

Muir and coworkers [16] take advantage of expressed protein ligation to create a novel phosphorylated protein that cannot be produced by other methods. The Csk kinase catalyzes the tyrosine phosphorylation of Src family members at a specific position on the carboxy-terminal tail, promoting an intramolecular association between the phosphotyrosine and the Src homology 2 (SH2) domain and repressing Src catalytic activity *in vivo*. Csk itself is related to the Src family but lacks the carboxy-terminal tail, and its activity is not regulated by phosphorylation. Muir and co-workers [16] asked whether the presence of an introduced phosphotyrosine at the carboxyl terminus of Csk would affect its activity or conformation. They used expressed protein ligation to attach a short tyrosine-containing peptide, either in the phosphorylated or unphosphorylated form, to the carboxyl terminus of Csk, such that the spacing between the tyrosine and the SH2 domain was identical to that in other Src family members. The technique also enabled a fluorescent tag to be incorporated into the semisynthetic protein to facilitate its detection. Indeed, phosphorylation of Csk, like that of other Src proteins, changes the protein conformation, suggesting that an intramolecular interaction forms between the SH2 domain and the introduced phosphotyrosine. Interestingly, the protein is approximately fivefold more active in the phosphorylated state, unlike Src, which is repressed by carboxy-terminal phosphorylation.

Expressed protein ligation has also been used to study protein–protein interactions. A conserved carboxy-terminal region from the *Escherichia coli* σ^{70} subunit of RNA polymerase has been implicated in contacting the –35 region of transcriptional promoters and in interacting with other subunits within the polymerase holoenzyme. Severinov and Muir [17] show that a truncated σ^{70} protein missing this region is transcriptionally inactive, but a semisynthetic protein made by restoring the carboxy-terminal 34 residues using a synthetic peptide is nearly as active as wild-type σ^{70}. Interestingly, insertion of an added cysteine residue at the ligation junction had no effect on activity, but whether this will be true in general will depend on the specific protein and the insertion site. In future studies, the interaction of the σ^{70} carboxy-terminal region with the other polymerase subunits can be examined by introducing cross-linking reagents or fluorescent probes to measure intersubunit distances.

Xu and coworkers [18] have shown that two proteins, RNase A and the restriction endonuclease *Hpa*I, both of which can be potentially difficult to express *in vivo* because of cytotoxic effects, can be readily produced using expressed protein ligation. This was done by introducing thioesters at the carboxy-terminal ends of truncated, inactive forms of the two proteins, which were then restored to the full-length RNase A and *Hpa*I proteins by ligation of 15- and 31-residue peptides, respectively. Ligation of the peptides is very efficient in both cases (> 80%). The results illustrate that refolding of the ligated products might be required to restore activity, however, and this is not always possible. The ligated RNase A was initially inactive, but the protein could be refolded and activity comparable to that of native RNase A was recovered after renaturation. In the case of *Hpa*I, ~95–98% of the ligated protein was insoluble, but the protein could not be refolded. The small amount of protein that was soluble

was nearly as active as the native endonuclease. Clearly, the amounts of recovered activity will be dependent on the protein and the splice-site position.

Conclusions

Rapid progress has been made in developing practical applications based on protein splicing. *Trans* splicing is now possible both *in vivo* and *in vitro*, and expressed protein ligation has been put into practice to introduce fluorescent and biotinylated tags into proteins, to produce otherwise cytotoxic proteins and to design novel modified proteins that could not have been engineered using other means. In the future, we could see this technique extended to allow synthetic peptides to be introduced into the interiors of proteins by performing sequential rounds of ligation. Exciting times lie ahead.

Acknowledgements

We thank Ming-Qun Xu and Xiang-Qin Liu for providing copies of their manuscripts prior to publication. This work was supported by a grant from the National Institutes of Health (GM50815).

References

1. Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. & Stevens, T.H. (1990). Protein splicing converts the yeast *TFP1* gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase. *Science* **250**, 651-657.
2. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990). Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H^+-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 6726-6733.
3. Belfort, M. & Roberts, R.J. (1997). Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* **25**, 3379-3388.
4. Perler, F.B., Xu, M.-Q. & Paulus, H. (1997). Protein splicing and autoproteolysis mechanisms. *Curr. Opin. Chem. Biol.* **1**, 292-299.
5. Gorbalenya, A.E. (1998). Non-canonical inteins. *Nucleic Acids Res.* **26**, 1741-1748.
6. Wu, H., Hu, Z. & Liu, X.-Q. (1998) Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **95**, 9226-9231.
7. Porter, J.A., *et al.*, & Beachy, P.A. (1996). Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21-34.
8. Duan, X., Gimble, F.S. & Quiocho, F.A. (1997). Crystal structure of PI-*Sce*I, a homing endonuclease with protein splicing activity. *Cell* **89**, 555-564.
9. Hall, T.M.T., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A. & Leahy, D.J. (1997). Crystal structure of a hedgehog autoprocessing domain: homology between hedgehog and self-splicing proteins. *Cell* **91**, 85-97.
10. Shingledecker, K., Jiang, S.-Q. & Paulus, H. (1998). Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: design of a minimal intein and of a *trans*-splicing system involving two intein fragments. *Gene* **207**, 187-195.
11. Wu, H., Xu, M.-Q. & Liu, X.-Q. (1998). Protein *trans*-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim. Biophys. Acta*, in press.
12. Mills, K.V., Lew, B.M., Jiang, S.-Q. & Paulus, H. (1998). Protein splicing in *trans* by purified N-and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proc. Natl Acad. Sci. USA* **95**, 3543-3548.
13. Southworth, M.W., Adam, E., Panne, D., Byer, R., Kautz, R. & Perler, F.B. (1998). Control of protein splicing by intein fragment reassembly. *EMBO J.* **17**, 918-926.
14. Lew, B.M., Mills, K.V. & Paulus, H. (1998). Protein splicing *in vitro* with a semisynthetic two-component minimal intein. *J. Biol. Chem.* **273**, 15887-15890.
15. Dawson, P.E., Muir, T.W., Clark-Lewis, I., & Kent, S.B. (1994). Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779.
16. Muir, T.W., Sondhi, D., & Cole, P.A. (1998). Expressed protein ligation: a general method for protein engineering. *Proc. Natl Acad. Sci. USA* **95**, 6705-6710.
17. Severinov, K. & Muir, T.W. (1998). Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J. Biol. Chem.* **273**, 16205-16209.
18. Evans, T.C., Jr., Benner, J. & Xu, M.-Q. (1998). Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.*, in press.
19. Chong, S., *et al.*, & Xu, M.-Q. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271-281.