Intein as a Novel Strategy for Protein Purification

Elahe Seyed Hosseini[1], Mehdi Zeinoddini[1], Hamed Haddad Kashani[2]*, Atefeh Ghoddusi[1]

[1]Research Center for Sciences and Biotechnology, Lavizan, Tehran. Iran.
[2]Anatomical Sciences Research Center, Kashan University of Medical Sciences, Kashan, Iran.

*** Corresponding author:** Hamed Haddad Kashani, Anatomical Sciences Research Center, Kashan University of Medical Sciences, Kashan, Iran. Email: Hamedir2010@gmail.com, Tel:00989137430153.

Abstract: Inteins are parts of proteins that cut themselves out of the whole protein entirely on their own. An important development in the area of recombinant protein purification has been the incorporation of self-cleaving protein elements into a variety of fusion-based purification systems. Inteins appear most frequently in Archaea, but they are found in organisms belonging to all three domains of life and in viral and phage proteins. Most inteins consist of two domains: One is involved in autocatalytic splicing, and the other is an endonuclease that is important in the spread of inteins. This review focuses on the evolution and technical application of inteins and only briefly summarizes recent advances in the study of the catalytic activities and structures of inteins. We further investigated the recent expanded applications of the Express Protein Ligation (EPL) technology in the fields of proteomics and bioimaging. This work is expected to provide a rough outline for the evaluation of these methods for large-scale bioprocessing of a variety of products.
[Elahe Seyed Hosseini, Mehdi Zeinoddini, Hamed Haddad Kashani. **Intein as a Novel Strategy for Protein Purification.** *Life Sci J* 2012;9(4):2187-2195] (ISSN: 1097-8135). http://www.lifesciencesite.com. 325

Key Words: Intein, Extein, Protein, Self- Splicing, ATPase, DNA

Introduction

Inteins (internal proteins) are genetic elements similar to self-splicing introns; however, inteins are transcribed and translated together with their host protein. Only at the protein level do the inteins excise themselves from the host protein. The two portions of the host protein separated by the intein are called exteins (external proteins) (1,2,3). During the splicing process the intein is excised, the two exteins are joined by a peptide bond, and the host protein assumes its normal folding and function. The first intein was discovered in 1987 when the carrot and *Neurospora crassa* vacuolar ATPases were compared with a putative $Ca2^{+}$ pumping ATPase. The latter had been isolated as a gene whose mutation made yeast resistant against the calmodulin antagonist trifluoperazine (4). The beginning and end of the encoded protein was very similar to the vacuolar ATPase subunits whose sequences had been submitted to the databanks at the same time. However, the central region of the putative calcium pump had no similarity to any known ATPase. Rather, this portion showed weak similarity to endonucleases. Anraku's lab (5) isolated the cDNA for the yeast vacuolar ATPase A-subunit and found the same sequence, including the central region, that had been earlier described as the trifluoperazine resistance gene (4). Surprisingly, denaturing polyacrylamide gel electrophoresis of the isolated protein demonstrated that the catalytic subunit of the functioning yeast V-ATPase had a molecular weight of only 70 kDa, as expected for a subunit without the insertion. Subsequently Kane et al. (6) showed that the insertion was still present in the mRNA, that the whole protein including the insertion was translated, and that the insertion spliced itself out of the protein during posttranslational processing. The concept that one gene determines one enzyme has long been a common principle for genetic information transfer (18, 19). In 1990, evidence that one gene produces two enzymes via protein splicing emerged from the structural study of the Saccharomyces cerevisiae VMA1 gene (20) followed by its expression analyses (20,21). The nascent 120-kDa translational product of VMA1 excises out the 50-kDa VDE site-specific endonuclease (VMA1-derived endonuclease (22) or VMA1 intein) and splices the N- and C-terminal exteins to form the mature 70-kDa catalytic subunit of the vacuolar ATPase. Since the first discovery of the VMA1 intein, more than 100 putative inteins have been identified in eubacteria, archaea, and eukaea species (23). Though only few have been proven to undergo protein splicing, their N- and C-terminal junctions are highly conserved (see the VMA1 sequence shown in Fig.1).

	N-Extein	Intein	C-Extein	
	1 284	362 738	1.071	
VMA1	M—NYSNSDAIIYVG**C**FAKG—T**H**E—VVVH**NC**GERGNEMAEVL—D			Splicing
XC-VDE-His	MIIYVG**C**FAKG—T**H**E—VVVH**NC**GHHHHHHGR			++
MIIYVG-VDE-CGER-C_t	MIIYVG**C**FAKG—T**H**E—VVVH**NC**GERGNEMAEVL—D			++
N_t-MIIYVG-VDE-CGER	M—NYSNSDAIIYVG**C**FAKG—T**H**E—VVVH**NC**GERG			++
XA-VDE	MIIYVG**A**FAKG—T**H**E—VVVH**NA**GERG			–
X10SAS	MSNSDAIIYVG**S**FAKG—T**N**E—VVVH**AS**GERGNEMAE			–
X10SSS	MSNSDAIIYVG**S**FAKG—T**N**E—VVVH**SS**GERGNEMAE			–
X10SNS	MSNSDAIIYVG**S**FAKG—T**N**E—VVVH**NS**GERGNEMAE			+
X10SNC	MSNSDAIIYVG**S**FAKG—T**N**E—VVVH**NC**GERGNEMAE			++
X10CNS	MSNSDAIIYVG**C**FAKG—T**N**E—VVVH**NS**GERGNEMAE			++
X10CNC	MSNSDAIIYVG**C**FAKG—T**N**E—VVVH**NC**GERGNEMAE			++

Figure 1. Sequences of the *VMA1* protein and recombinants. The N- and C-terminal sequences of the recombinants are aligned with the *VMA1* sequence of which the residues from Cys284 to Asn737 correspond to the VDE intein. Residues subjected to the mutation studies are indicated with bold characters. The N- and C-terminal side polypeptides are indicated with N_t and C_t, respectively. The replacements in the X10-series VDE recombinants are underlined. Splicing activity for each recombinant discussed in the manuscript is represented with a corresponding sign in the 'Splicing' column.

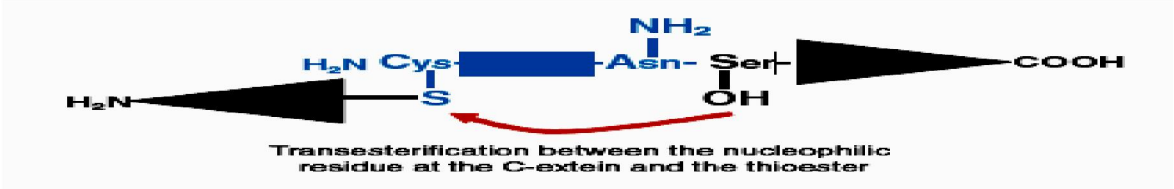
Mechanism of Protein Splicing

Substantial information about the chemical reactions involved in protein splicing is available (7, 8, 9, 10, 11, 12, 13, 14, 15) and was recently reviewed in (16). Briefly, protein splicing involves the following four steps.

Step 1 The amino-terminal splice junction of the intein is activated by an N-O or N-S shift that leads to an ester or thioester intermediate. As a result of this rearrangement, the N-extein binds to the oxygen of a serine or to the sulfur of a cysteine residue at the amino-terminal splice junction.

N-S acyl rearrangement of the peptide bond at the N-terminus

Step 2 Cleavage of the ester at the amino-terminal splice junction occurs through attack of a nucleophilic residue located at the carboxy-terminal splice junction. This transesterification results in a branched protein intermediate.



Transesterification between the nucleophilic residue at the C-extein and the thioester

Step 3 The cleavage proceeds through asparagine cyclization, which causes intein excision and splicing of the two exteins by an ester bond. Several inteins have glutamine rather than asparagine residues at their C-terminal end, suggesting that cleavage in these inteins might occur via an aminoglutarimide rather than an aminosuccinimide intermediate (16, 1 7).

Step 4 A spontaneous rearrangement results in formation of a peptide bond between the two exteins.

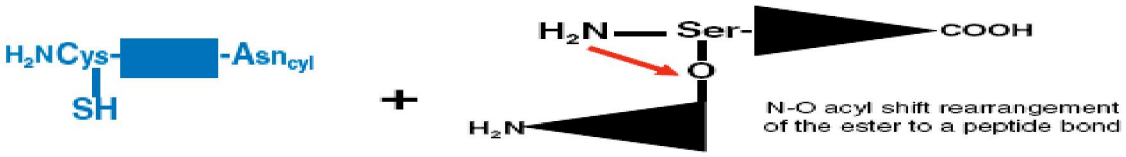


Figure 2. Structure of mini-inteins and large inteins

Inteins are classified into two groups, large and minimal (mini) (24). Large inteins contain a homing endonuclease domain that is absent in mini-inteins. Homing endonucleases are site-specific, double-strand DNA endonucleases that promote the lateral transfer between genomes of their own coding region with flanking sequences, in a recombination-dependent process known as "homing."Usually, homing endonucleases are licing domain, and a central endonuclease domain. Splicing-efficient mini-inteins have been engineered from large inteins by deleting the central endonuclease domain, demonstrating that the endonuclease domain is not involved in protein splicing (27,28,29). The splicing domain is split by the endonuclease domain into N- and C-terminal subdomains, encoded by an open reading frame within an intron or intein (25,26). Large inteins are bifunctional proteins, with a protein splicing which contain conserved blocks of amino acids, with blocks A, N2, B, andN4 in the N-terminal sub domain, and blocks G and F in the C-terminal subdomain (30, 31) These domains can also be identified in mini-inteins (Fig. 3). The three-dimensional structures of naturally occurring mini-inteins and engineered mini-inteins reveal that the N- and C-terminal splicing domains form a common horseshoe-like 12-β-strand scaffold termed the Hedgehog. Intein (HINT) module (32,33,34,35,36,37). All known inteins share a low degree of sequence similarity, with conserved residues only at the N- and C-termini. Most inteins begin with Ser or Cys and end in His-Asn, or in His-Gln. The first amino acid of the C-extein is an invariant Ser, Thr, or Cys, but the residue preceding the intein at the N-extein is not conserved (38). However, residues proximal to the intein-splicing junction at both the N- and C-terminal exteins were recently found to accelerate or attenuate protein splicing (39).

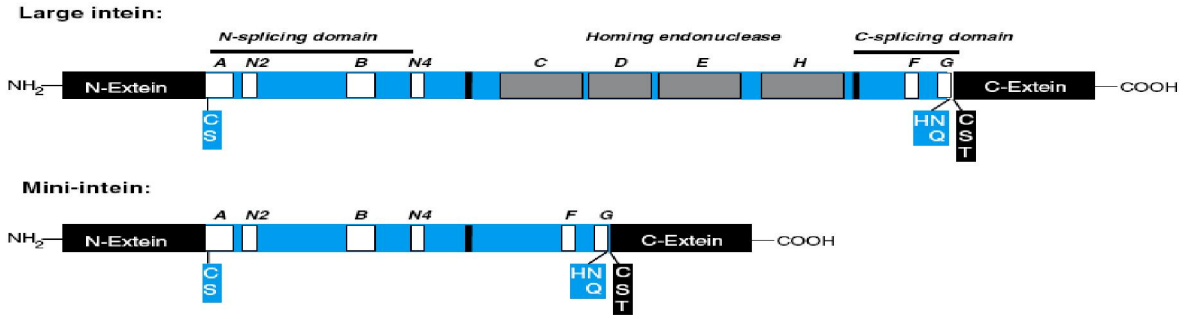


Figure 3: Structure of large and mini-inteins. Conserved elements in a large intein and mini-intein are indicated. The white and grey areas A, N2, B, N4, C, D, E, H, F, and G are conserved intein motifs identified by Pietrokovski (40) and Perler et al. (41). The exteins are illustrated in black and the intein sequence in blue.

Application of intein in biotechnology

1) Self-circularization by inteins

Cyclic antibiotics display an increased activity and stability in comparison to their linear analogs. Inteins play a growing role in the production of cyclic peptides through the aforementioned IPL technique (42). The protein of interest is fused through its C-terminus, to the N-terminus of an intein, in which the C-terminal Asn has been mutated to be incapable of cleaving the C-terminal binding tag. The N-terminus of the target protein is altered so the second residue after the Met is a Cys. After purification of the target protein, the Nterminal Met is removed using methionyl-aminopeptidase from E. coli, resulting in an N-terminal Cys residue (43,44). After purification, including elution from the column by thiol induced N-terminal cleavage of the intein, the linear peptide contains a C-terminal thioester and a Cys at its N-terminus that can react to form a new peptide bond. The introduction of a non-native linker sequence improves cleavage efficiency, but also has the potential to interfere with the biological activity of the cyclic protein. Another problem is the possibility of polymerization instead of cyclization by activated peptides Split inteins have also been applied in the generation of cyclic proteins and peptides. The very timely and elegant Split Intein-mediated Ciruclar Ligation Of Peptides a ProteinS (SICLOPPS) system uses the naturally split Synechocystis sp. Ssp DnaE intein, which is fused in a rearranged order (IC–target protein–IN), allowing the efficient cyclization of the target protein by reconstitution of the Ssp DnaE intein. Using this method, it was possible to generate cyclic peptides that are short as eight amino acid residues. SICLOPPS has been used in inhibitor studies for the rapid synthesis of very large cyclic peptide libraries that are superior to the traditional chemically generated libraries, and which can be screened in vivo for new potent therapeutic drugs (45, 44). In recent years, SICLOPPS has been impressively used, for instance, to identify several inhibitors for the dimerization of ribonucleotide reductase and 5- aminoimidazole-4-carboxyamide-ribotide transformylase (46).

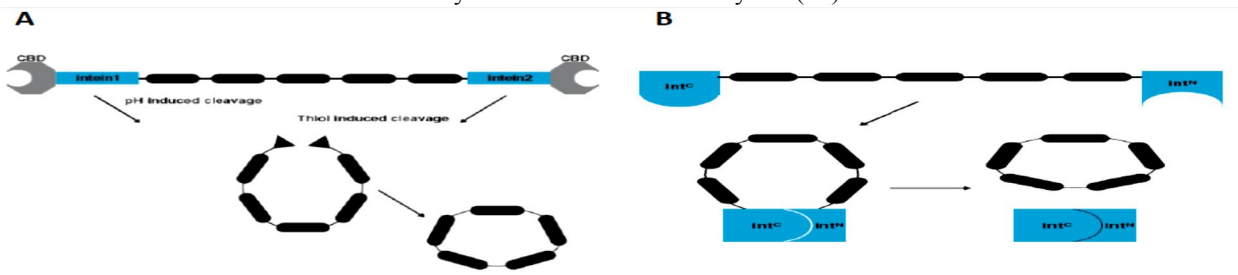


Figure 4. **(A)**Self-circularization of peptides. a Self-circularization of proteins using the TWIN system. A target protein is embedded between two intein sequences, which are modified for N- or C-terminal cleavage, respectively. The inducible splicing reaction of the inteins leads to the generation of an activated thioester residue and an N-terminal Cys for the spontaneous circularization of the linear peptide. **(B)**Utilization of the Split Intein-mediated Ciruclar Ligation of Peptides a Protein S (SICLOPPS) also enables the circularization of peptides. In this system, the order of the naturally split Synechocystis sp. Ssp DnaE intein is inverted (IC–target protein–IN), and the reconstitution of the SsDnaE intein allows the efficient cyclization of the target protein.

2) Intein-Mediated, *In Vitro* and *In Vivo* Protein Modifications with Small Molecules

Native chemical ligation was developed for the synthesis of peptides by condensation of their unprotected segments (47). This technique requires a synthetic peptide with a Cterminal thioester and another with an N-terminal cysteine residue to yield a native peptide bond at the site of ligation. Though useful, this approach was limited to the synthesis of small proteins of 100-200 residues in size, as multiple ligations have proven to be technically demanding. By combining the intein-mediated protein splicing with the chemistry of native chemical ligation, it is now possible to engineer unnatural proteins of, in principle at least, unlimited sizes. This new approach is known as the Expressed Protein Ligation or intein-mediated protein ligation (EPL or IPL) (48). Briefly, it refers to the use of an Intein-based Protein Ligation system to generate a protein thioester from the corresponding protein-intein fusion. Unique functionalities can then be incorporated into the proteins using the chemistry of native chemical ligation. Over the years, intein mediated protein ligation has become an extremely powerful platform for the chemical semi-synthesis and engineering of proteins. For instance, the technique has been used for the incorporation of non coded amino acids into protein sequences (49), the isolation of cytotoxic proteins (49), and the synthesis of the 600-amino acid N-terminal segment of the 70 subunit of *Escherichia. coli* RNA polymerase(49), just to name a few. Tran-splicing inteins (50), in which functionally mature intein was split into two fragments and only regained their activity upon reconstitution of the two fragments, have been employed in protein semi-synthesis(50), and for the study of protein structure/function relationship by segmental isotopic labeling of proteins for NMR analysis (50). At least two independent groups had reported the use of split inteins for the synthesis of cyclized proteins and peptides (51). Recently, the use of split inteins has been applied to the study of protein-protein interactions and the translocation of proteins in living cells (51). For a more in-depth coverage of the Express Protein Ligation and its applications, readers are directed to the most recent review by Muir (48). Biotin-avidin system is an interesting and highly exploited system for a plethora of diverse applications in modern biology. The success of this system would depend on the various approaches available for the incorporation of the biotin moiety into a protein of interest. Traditionally, protein biotinylation has been achieved using simple bioconjugate techniques with biotin-containing reactive chemicals (53). This typically results in the non specific biotinylation and, in many cases, the inevitable inactivation of the target protein. Alternative techniques for the site-specific biotinylation of proteins have been developed (54, 55). Cronan *et al.* identified a set of peptide sequences, which upon fusion to a target protein could be biotinylated by the enzyme biotin ligase. The biotin ligase works by covalently attaching the biotin moiety to a specific lysine residue in the peptide sequence (54). This thus provides a site-specific approach for protein biotinylation, both *in vitro* and *in vivo*, although the process was often inefficient and toxic to the host cell when performed *in vivo* (55). To overcome some of the limitations faced in current protein biotinylation techniques, the use of intein-mediated protein ligation for the site-specific biotinylation of proteins were investigated (56). The protein of interest was fused through its C-terminus to an intein, which had its other terminus fused to a chitin-binding domain (56,56). Upon expression in the host cell, the cell lysate containing the fusion protein was first loaded onto a column packed with chitin beads followed by the addition of a thiol-cleaving reagent (e.g. cysteine-biotin; inset in Figure 5).

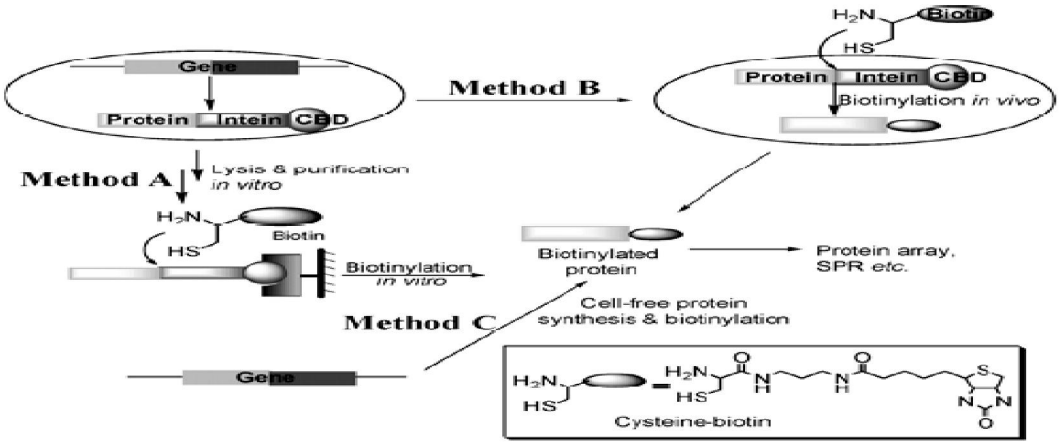


Figure 5. Three intein-mediated protein biotinylation strategies. (Method A) *In vitro* biotinylation of column- bound proteins; (Method B) *In vivo* biotinylation in live cells; (Method C) Cell-free biotinylation of proteins.

The fusion protein underwent self-cleavage, catalyzed by the fused intein, and resulted in the protein having a reactive thioester group at the C-terminal. At this stage, the thiol side chain from the addition of cysteine-biotin attacked the thioester functionality and the resulting thioester-linked intermediate underwent spontaneous rearrangement, affording a native peptide bond to generate proteins that were site-specifically biotinylated at their C-termini (Figure 5; Method A). This strategy is simple yet efficient and capable of biotinylating various proteins from different biological sources. In addition, this strategy could be readily adopted in a 96-well format, enabling highthroughput generation of a large number of biotinylated proteins. Next, this strategy extended to *in vivo* biotinylation of proteins in both bacterial and mammalian systems (Figure 5; Method B). The simple addition of the cell-permeable cysteine-biotin probe to the growth media followed by further incubation of the cells resulted in substantial biotinylation of the proteins inside the cells. Further optimizations of the cell growth and *in vivo* biotinylation conditions led to an increase in the level of protein biotinylation in the cells. For the bacterial system, the only other band observed on the western blot was acetyl-CoA carboxylase, an endogenous biotinylated protein found in bacterial cells. Similarly for the mammalian system, the three other biotinylated proteins observed on the western blots, besides the target protein, were the endogenous biotinylated proteins. The efficiency of intein-mediated protein biotinylation, both *in vitro* and *in vivo*, depends greatly on the intein fused to target protein (56). A cell free system has many potential advantages over traditional *in vivo* recombinant protein expression in cells (57). Cellular toxicities due to overexpression of certain proteins, possible degradation by endogenous proteases and formation of inclusion bodies of proteins could potentially be avoided all together in a cell free system. It should be noted however, although the cell free biotinylation approach seems the most simple out of the three strategies presented in Figure 5, the efficiency of protein expression and subsequent biotinylation depended on a few other factors including the DNA quality and concentration, the nature of the target protein and etc. The utility of the above-described site-specific biotinylation of proteins is in the field of protein arrays. Array-based technologies are one of the most promising technologies developed in the post-genomic era to study proteins in a high-throughput fashion (58). One of the key aspects in generating a uniform functional protein array is the ability to immobilize proteins onto glass slides and at the same time ensuring that their native activity is not lost. This is because proteins are "delicate" they may unfold and lose their activity if not properly attached to a suitable surface. While a variety of approaches have been developed which allow site-specific immobilization of small molecules and peptides in a microarray (59), relatively few methods are available for protein immobilization (58, 60-62). In most cases, these approaches are non-specific, resulting in randomly orientated protein immobilization (58,60). Uniform site-specific immobilizations of functional proteins was first reported by Zhu *et al.*, who successfully created a functional "yeast proteome array" by attaching thousands of (His) 6-tagged proteins onto Ni-NTA-coated glass slides (60). However, the noncovalent (His) 6-Ni-NTA interaction is not very strong, often susceptible to interference by commonly used chemicals and salts, making this strategy useful to only a handful of down-stream protein array applications (56, 61, 62,63). Mrksich and co-workers captured cultinase-fused proteins onto glass surfaces coated with a phosphonate ligand, achieving site-specific and covalent immobilization of the proteins (61). Scientists successfully developed a site specific method to covalently immobilize hAGT-fused proteins onto modified glass surfaces (61). Also they used Sfp phosphopantetheinyl transferase to mediate site-specific covalent immobilization of target proteins fused to the peptide carrier protein (PCP) excised from a nonribosomal peptide synthetase (NRPS) (61). Recently, scientists described the use of EPL to express functionally active proteins which possess a C-terminal thioester handle, and subsequently immobilized them to a cysteine-modified glass slide, generating the corresponding protein array (61).

Summary and Future Outlook

The site-specific modification of proteins represents a powerful approach by which one can manipulate/ control the protein activities and distribution, thereby studying and understanding the biological function of proteins in the cell. Intein-mediated protein ligation is one such tool that allows the specific incorporation of novel functionalities into proteins. Both strategies were applied to the creation of protein arrays using biotin-avidin system and the chemistry of native chemical ligation, respectively. Scientists also extended the EPL to the field of bioimaging by synthesizing different cell-permeable small molecules probes that are able to site-specifically label N-terminal cysteine containing proteins expressed inside a living cell. Current efforts are looking at other forms of proteins that undergo auto-proteolysis. The family of hedgehog proteins found in animals is one example (64). These proteins are responsible for the patterning of a variety of structure in animal embryogenesis. Hedgehog proteins were found to undergo a self-cleavage reaction, which

was mediated by its carboxyl-terminal end to generate reactive intermediates required for protein ligation. Hence the study of hedgehog proteins might open up another possibility for mammalian-based protein modifications/engineering.

Currently, there are very few enzyme-based approaches for protein ligation (65, 66). One is subtiligase, an engineered subtilisin that have been shown to catalyze the ligation between peptide fragments through a process known as reverse proteolysis (65). Another is a transpeptidase found in the cell envelope of many Gram-positive bacteria Sortase (66). Sortase recognizes an LPXTG motif and cleaves between the threonine and glycine residue pentaglycine located on the peptidoglycan of the cell wall (66).

***Corresponding author:** Hamed Haddad Kashani, Anatomical Sciences Research Center, Kashan University of Medical Sciences, Kashan, Iran. Email: Hamedir2010@gmail.com, Tel:00989137430153.

References:

1. Cooper AA, Stevens TH. Protein splicing: self-splicing of genetically mobile elements at the protein level. Trends Biochem. Sci. 1995. 20:351–56.
2. Dujon B.. Group I introns as mobile genetic elements: facts and mechanistic speculations a review. Gene1989. 82:91–114.
3. Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, et al. Protein splicing elements: inteins and exteins a definition of terms and recommended nomenclature. Nucleic Acids Res. 1994. 22:1125–27.
4. Shih CK, Wagner R, Feinstein S, Kanik- Ennulat C, Neff N. A dominant trifluoperazine resistance gene from Saccharomyces cerevisiae has homology with F0F1 ATP synthase and confers calciumsensitive growth. Mol. Cell. Biol.1988. 8:3094–103.
5. Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)- translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae. J. Biol. Chem. 1990. 265: 6726–33.
6. Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science. 1990. 250:651–57.
7. Chong S, Montello GE, Zhang A, Cantor EJ, Liao W, et al. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. Nucleic Acids Res. 1998. 26:5109–15.
8. Cooper AA, Chen YJ, Lindorfer MA, Stevens TH. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. EMBO J. 1993. 12: 2575–83.
9. Mathys S, Evans TC, Chute IC, Wu H, Chong S, et al. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene. 1999. 231:1–13.
10. Paulus H. Protein splicing and related forms of protein autoprocessing. Annu. Rev. Biochem. 2000. 69:447–96.
11. Perler FB, Xu MQ, Paulus H. Protein splicing and autoproteolysis mechanisms. Curr. Opin. Chem. Biol. 1997. 1:292– 99.
12. Poland BW, Xu MQ, Quiocho FA. Structural insights into the protein splicing mechanism of PI-sceI. J. Biol. Chem. 2000. 275:16408–1.
13. Shao Y, Paulus H. Protein splicing: estimation of the rate of O-N and S-N acyl rearrangements, the last step of the splicing process. J. Pept. Res. 1997. 50:193–98.
14. Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB. Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. Biotechniques. 1999. 27: 110–4, 116, 118- 20.
15. Xu MQ, Perler FB. The mechanism of protein splicing and its modulation by mutation. EMBO J. 1996. 15:5146–53.
16. Paulus H. Protein splicing and related forms of protein autoprocessing. Annu. Rev. Biochem. 2000. 69:447–96.
17. Pietrokovski S. Identification of a virus intein and a possible variation in the protein-splicing reaction. Curr. Biol. 1998.8:R634–35.
18. Beadle, G. W. The genetic control of biochemical reactions. Harvey Lectures. 1945. 40, 179 – 194.
19. Horowitz, N. H. The one gene-one enzyme hypothesis. Genetics. 1948. 33, 612 – 613.
20. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H+-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae. J. Biol. Chem. 1990. 265, 6726 – 6733.
21. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H+-adenosine triphosphatase. Science. (1990). 250, 651 – 657.

22. Gimble, F. S., and Thorner, J. Homing of a DNA endonuclease gene by meiotic gene conversion in Saccharomyces cerevisiae. Nature. (1992) 357, 301 – 306.
23. Perler, F. B. InBase, the Intein Database. Nucleic Acids Res. 2000 January 1; 28(1): 344–345.
24. Liu XQ. Protein-splicing intein: genetic mobility, origin, and evolution. Ann Rev Genet. 2000. 34:61–76.
25. Belfort M, Derbyshire V, Stoddard BL, Wood DW. Homing endonucleases and inteins. Springer, Berlin Heidelberg New York. 2005 pp. 273 –292.
26. Chevalier BS, Stoddard BL. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucl Acids Res. 2001. 29:3757–3774.
27. Chong S, Xu MQ. Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs. J Biol Chem. 1997. 272:15587–155890.
28. Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M. Genetic definition of a protein-splicing domain:functional mini-inteins support structure predictions and a model for intein evolution. Proc Natl Acad Sci U S A.1997. 94:11466–11471.
29. Shingledecker K, Jiang SQ, Paulus H. Molecular dissection of the Mycobacterium tuberculosis RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments. Gene. 1998. 207:187–195.
30. Perler FB, Olsen GJ, Adam E. Compilation and analysis of intein sequences. Nucl Acids Res. 1997. 25:1087–1093.
31. Pietrokovski S. Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. Protein Sci.1994. 3:2340–2350.
32. Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, Lesage G, Rao Z. Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagines during protein splicing. J Biol Chem. 2003. 278:39133–39142.
33. Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ. Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. Cell. 1997. 91:85–97.
34. Klabunde T, Sharma S, Telenti A, Jacobs WRJ, Sacchettini JC. Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing. Nat Struct Biol. 1998. 5:31–36.
35. Koonin EV. A protein splice-junction motif in hedgehog family proteins. Trends Biochem Sci. 1995. 20:141–142.
36. Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z . Crystal structures of an intein from the split dnaE gene of Synechocystis sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc Ion inhibition of protein splicing. J Mol Biol. 2005. 353:1093–1105.
37. Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V. Crystallographic and mutational studies of Mycobacterium tuberculosis recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. J Mol Biol. 2007. 367:162–173.
38. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. Modulation of intein activity by its neighboring extein substrates. Proc Natl Acad Sci U S A. 2009. 106:11005–11010.
39. Perler FB . InBase: the Intein Database. Nucl Acids Res. 2002. 30:383–384.
40. Pietrokovski S. Identification of a virus intein and a possible variation in the protein-splicing reaction. Curr Biol.1998. 8:R634–R635.
41. Perler FB. Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. Cell. 1998.92.
42. Evans TC Jr, Benner J, Xu MQ. The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. J Biol Chem.1999. 274:18359–18363.
43. Sancheti H, Camarero JA. “Splicing up” drug discovery. Cellbased expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. Adv Drug Deliv Rev. 2009. 61:908–917.
44. Tavassoli A, Naumann TA, Benkovic SJ. Production of cyclic proteins and peptides. In: Belfort M, Derbyshire V, Stoddard BL, Wood DW (eds) Homing endonucleases and inteins, vol. 16. Springer, Berlin Heidelberg New York,. 2005. pp 293–305.
45. Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ. Production of cyclic peptides and proteins in vivo. Proc Natl Acad Sci U S A. 1999.96:13638–13643.
46. Cheriyan M, Perler FB. Protein splicing: a versatile tool for drug discovery. Adv Drug Deliv Rev. 2009. 61:899–907:1–4.
47. Dawson, P.E.; Kent, S.B.H. Synthesis of native proteins by chemical ligation. Ann. Rev. Biochem., 2000. 69, 923- 960.
48. Muir, T.W. Semisynthesis of proteins by expressed protein ligation. Ann. Rev. Biochem., 2003. 72, 249-289.

49. Ayers, B.; Blaschke, U.K; Camarero, J.A.; Cotton, G.J.; Holford, M.; Muir, T.W. Mapping the molecular interface between the sigma. Biopolymers, 1999. 51, 343-354.
50. Paulus, H. Protein Splicing and Related Forms of Protein Autoprocessing. Ann. Rev. Biochem., 2000. 69, 447-496.
51. Scott, C.P.; Abel-Santos, E.; Wall, M.; Wahnon, D.C.; Benkovic, Production of cyclic peptides and proteins in vivo. S.J. Proc. Natl. Acad. Sci. USA, 1999. 96, 13638–13643.
52. Wilchek, M.; Bayer, E.A. Introduction to avidin-biotin technology. Methods Enzymol., 1990. 184, 5-13.
53. Hermanson, G.T. Bioconjugate techniques, Academic Press, San Diego, CA. 1996, 785 pages.
54. Cronan, J.E.; Reed, K.E. Biotinylation of proteins in vivo: a useful · posttranslational modification for protein analysis. Methods Enzymol., 2000. 27, 440-458.
55. Cull, M.G.; Schatz, P.J. Biotinylation of proteins in vitro and in vivo using small peptide tags. Methods Enzymol., 2000. 26, 430-440.
56. Lue, R.Y.P.; Chen, G.Y.J.; Hu, Y.; Zhu, Q.; Yao, S.Q. Versatile protein biotinylation strategies for potential high-throughput proteomics. J. Am. Chem. Soc., 2004. 126, 1055-1062.
57. He, M.Y.; Taussig, M.J. Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation. Nucleic Acids Res., 2001. 29, e73.
58. MacBeath, G.; Schreiber, S.L. Printing proteins as microarrays for high throughput function determination. Science, 2000, 289, 1760-1763.
59. Walsh, D.P.; Chang, Y.T. Recent Advances in Small Molecule Microarrays, Applications and Technology.Comb. Chem. High Throughput Screening, 2004. 7, 557-564.
60. Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R.A.; Gerstein, M.; Snyder, M. Global analysis of protein activities using proteome chips. Science, 2001, 293, 2101-2105.
61. Mrksich, M. Hodneland, C.D.; Lee, Y.S.; Min, D.H. Selective immobilization of proteins to self-assembled monolayers presenting active site-directed capture ligands. Proc. Natl. Acad. Sci. USA, 2002. 99, 5048-5052.
62. Perler, F.B. Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. Cell, 1998. 92, 1-4.
63. Xu, M.Q.; Evans, T.C. Intein-Mediated Ligation and Cyclization of Expressed Proteins. Methods, 2001. 24, 257-277.
64. Mann, R.K., Beachy, P.A. Cholesterol modification of proteins. Biochim. Biophy. Acta., 2000. 1529, 188-202.
65. Jackson, D.Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J.A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues. Science, 1994. 266, 243-247.
66. Phan, A.; Zdanov, A.G.; Evdokimov, J.E.; Tropea, H.K.; Peters III, R.B.; Kapust, M.; Li, A.; Wlodawer, D.S.; Waugh, Structural basis for the substrate specificity of tobacco etch virus protease. J. Biol. Chem., 2002. 277, 50564-50572.
67. Yeo, S.Y.D.; Srinivasan, R.; Chen, G.Y.J.; Yao, S.Q. Expanded utilities of the native chemical ligation reaction. Chem. Eur. J., 2004. 10, 4664-4672.

10/10/2012