

蛋白质剪接及在蛋白质工程中的应用

赫冬梅^{1,2}, 钱凯先¹, 沈桂芳^{1,2}

(1. 浙江大学生命科学院, 杭州 310027; 2. 中国农业科学院生物技术中心, 北京 100081)

摘要:蛋白质剪接是蛋白质内含肽介导的, 一种在蛋白质水平上翻译后的加工过程, 它由一系列分子内的剪切—连接反应组成。蛋白质内含肽是一个蛋白质前体中的多肽序列, 可以催化自身从蛋白质前体中断裂, 使两侧的蛋白质外显肽连接成成熟的蛋白质。蛋白质内含肽的发现, 不仅丰富了遗传信息翻译后加工的理论, 在实践中也有广泛的应用前景。

关键词:蛋白质剪接; 蛋白质内含肽; 蛋白质外显肽

中图分类号: Q51

文献标识码: A

文章编号: 0253-9772(2004)02-0249-04

Protein Splicing and Its Application in Protein Engineering

HE Dong-Mei^{1,2}, QIAN Kai-Xian¹, SHEN Gui-Fang^{1,2}

(1. College of Biology, Zhejiang University, Hangzhou 310027, China;

2. Biotechnology Research Institute, CAAS, Beijing 100081, China)

Abstract: Protein splicing, which is an intein mediated posttranslational processing, involves a series of intramolecular cleavage-ligation reactions. Intein is an intervening polypeptide which can catalytic self-cleavage from a pre-protein accompanied by the concomitant joining of the two flanking polypeptides (the extein) through a peptide bond. Protein splicing not only enriches genetic theory of posttranslational processing, but also has a wide application prospect.

Key words: protein splicing; intein; extein

自 20 世纪 90 年代初发现蛋白质剪接现象以来, 就引起了有关专家的关注。蛋白质剪接 (protein splicing) 是在蛋白质内含肽 (intein) 的自我催化作用下, 从翻译后的蛋白质前体中切除蛋白质内含肽, 同时, 两侧的蛋白质外显肽 (extein) 通过连接反应, 形成一个新的具有生物活性的蛋白质^[1,2]。蛋白质剪接作用与 mRNA 水平上的修饰作用有相似之处, 但它既不需要辅酶也不需要辅因子。蛋白质内含肽是蛋白质剪接中的主要遗传元件^[3], 它的编码基因插入到编码前体蛋白质的基因序列中, 和两侧的序列一起转录成 mRNA, 并翻译成多肽链, 之后, 这段担负自我催化作用的内含肽被切掉, 蛋白质外显肽连接成成熟蛋白质。蛋白质内含肽 (intein) 和 RNA 内含子 (intron) 一样, 都属于可移动的遗传元件, 通过在 RNA 或蛋白质水平上的自我剪切而形成成熟的

mRNA 和有功能的蛋白质^[4]。与 intron 不同的是, 蛋白质内含肽和宿主蛋白一起转录和翻译成前体蛋白 (如图 1)。蛋白质内含肽主要存在于单细胞生物中, 如: 嗜热古细菌、杆菌及酵母中^[5]。迄今已经纯化了近百种蛋白质内含肽 (<http://www.neb.com/neb/intein.html>), 并围绕其结构、机理、应用等进行了一系列研究^[6,7]。

1 蛋白质的剪接

自从 1990 年在啤酒酵母 (*S. cerevisiae*) ATP 中发现第一个蛋白质内含肽 (Sec VMA) 后, 其数目不断增加^[1]。分析蛋白质内含肽的序列特征, 对理解它的结构与功能的关系非常必要。

已经发现的近百种蛋白质内含肽, 大小均在 134~608

收稿日期: 2003-05-30; 修回日期: 2003-10-08

基金项目: 农业部 948 项目资助 (编号: 991020) [Subsidized by 948 item Agricultural Department of China (No. 991020)]

作者简介: 赫冬梅 (1972-), 女, 浙江大学在读博士研究生, 专业方向: 细胞及分子生物学

通讯作者: 沈桂芳 (1932-), 教授, 分子生物学和遗传工程, Tel: 010-68919854, E-mail: dmh11617@sohu.com

个氨基酸之间。其中 24 种蛋白质内含肽具有下列典型的蛋白质剪接的结构特点^[8]:(1)在蛋白质编码序列的一定位置具有插入序列,而其同源蛋白则不存在;(2)蛋白质内含肽的 N 末端有保守的 Cys 或 Ser 残基,而蛋白质外显肽在剪接处则具有 Cys, Ser 或 Thr 残基;(3)蛋白质内含肽的 C 末端具有特征性短序列,如: His-Asn。序列比较分析表明^[9]: 已知的蛋白质内含肽都具有双功能(如图 2): 即 N 末端的 100~150 个氨基酸残基和 C 末端的 50 个氨基酸残基组成蛋白质剪接域;中间的 200~250 个氨基酸残基具有自导引核酸内切酶活性。蛋白质内含肽的两个剪切域约 150~200 个氨基酸,可以激活内含肽的 N 末和 C 末端剪接处肽链的断裂,并通过形成一个新的肽键连接成为具有活性的蛋白质分子。

还有一些蛋白质内含肽只有 130~200 个氨基酸,不含自导引核酸内切酶。

研究表明^[10]: 蛋白质内含肽不仅可以介导蛋白质水平上的加工,而且由于蛋白质内含肽的某些特定区域,在结构和功能上与核酸内切酶相似,还可以进行内含肽自身的转移,但这一区域不具有蛋白质剪接功能。蛋白质内含肽作为可移动因子,可通过“内含肽寻靶”进行转导。内含肽寻靶(homing)是指内含肽这段插入序列,转移到没有这一序列的同源蛋白的等位基因的过程。自导引核酸内切酶可以识别、切割不含蛋白质内含子编码序列的同源蛋白质的等位基因,并利用同源重组,将编码蛋白质内含肽的基因拷贝转移到不含蛋白质内含肽的等位基因上^[11]。

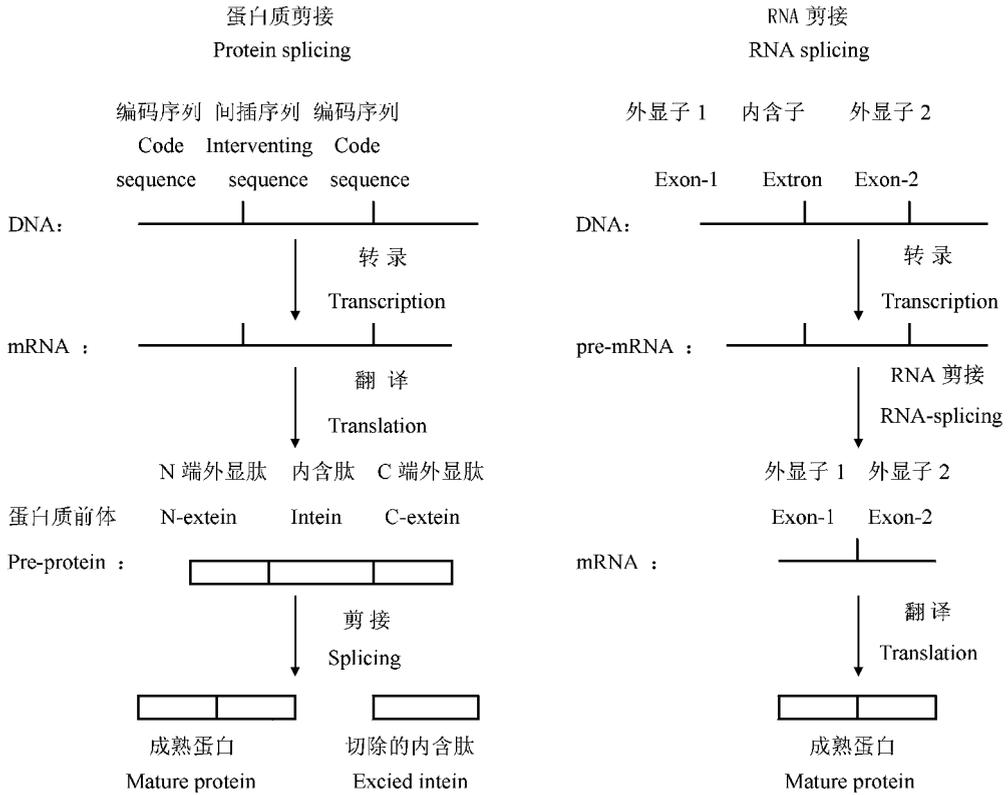


图 1 蛋白剪接和 RNA 剪接的比较

Fig. 1 Comparison of Protein splicing and RNA splicing

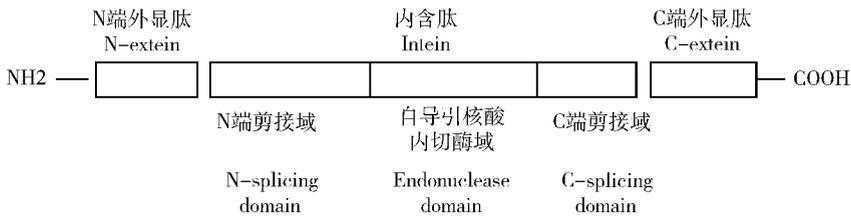


图 2 蛋白质内含肽的结构

Fig. 2 Structure of intein

序列统计模型和结构晶体学分析表明,蛋白质内含肽有两个功能结构域,在结构和功能上是独立的,不同来源的两个部分可以融合成新的蛋白质内含肽^[12],这为重组蛋白质内含肽提供了可能。

2 蛋白质剪接的机理

研究表明,蛋白质剪接反应,发生在蛋白质内含肽的保守残基形成的活性中心内,包括分子内的转换、中间产物的形成、Asp 的环化、肽键的断裂和形成等四个步骤。其中,肽键的断裂和形成是蛋白质剪接的关键反应,而 *N-O* 或 *N-S* 酯酰基的形成也是蛋白质剪接的必要条件。为了深入了解蛋白质剪接的机理,Larke 等通过人工诱导的基因突变,抑制蛋白质剪接,分别对前体、中间产物及终产物加以研究,提出了蛋白质断裂和连接的连续反应机理的假说^[13]。

2.1 *N-O* 或 *N-S* 酯酰基的重组

即蛋白质内含肽 *N* 末端剪切点的 Ser/Cys 残基侧链的 *N-O* 或 *N-S* 发生转酯酰基反应,形成了一个线性中间产物——酯类。基因突变和化学断裂表明:*N-O* 或 *N-S* 酯酰基的重组是内含肽 *N* 末端断裂作用的关键,Ser 残基的取代不发生转酯作用,*N* 末端也不断裂,表明 Ser 对启动剪接过程和 *N* 末端的断裂是必需的^[14]。

2.2 转酯作用

转酯作用需要具有亲核残基的参与,蛋白质内含肽 *C* 端剪切点的 Ser/Thr/Cys 残基亲核攻击 *N* 末端的(硫)酯键,使上游外显肽从蛋白质内含肽的 *N* 末端分离,转到下游外显肽的 Ser/Thr/Cys 的侧链上,形成分支酯类中间产物。

2.3 Asp 残基的环化

多肽或蛋白质内含肽中 *C* 末端的 Asp 残基环化形成氨基琥珀酰亚胺,随后分支中间产物断裂,具有氨基琥珀酰亚胺的内含肽被切除。在这个反应中,Asp 的 β -酰胺键发生脱氨作用。

2.4 *S-N* 或 *O-N* 酰基的重组

是一个自发反应。随着 Asp 的环化和内含肽与 *C* 端外显肽之间肽键的断裂,内含肽被切除,外显肽相连,但切掉的内含肽的氨基琥珀酰亚胺还要发生自发羟基化,两侧的外显肽之间形成稳定的肽键,才完成蛋白质的剪接。

前 3 步反应必须在内含肽的“催化”作用下发生,非自发性的;最后一步是瞬时的、自发的。从这个意义上讲,蛋白质内含肽是一种狭义上的“酶”,其作用底物是两侧剪接处的保守氨基酸残基。

3 蛋白质剪接的调节及其应用

如前所述,蛋白质剪接的前 3 步反应可独立发生,因此对剪切处的保守氨基酸的定位及功能分析,使蛋白质剪接的调节成为可能。用重组的蛋白质内含肽可以调控蛋白质的剪接过程,并广泛用于蛋白质的合成、分离纯化、标记、检测

等。

3.1 快速分离、纯化目的蛋白

Chong S R 等^[15,16]把可自我裂解的亲标记,连在目标蛋白的 *C* 末或 *N* 末端,用亲和层析可以实现一步纯化蛋白质。早期使用的是 *S cerevisiae* VMA 内含肽,先用 Ala 取代蛋白质内含肽 *C* 末端的 Asp 残基,再连接上几丁质(chitin)作为 *C* 末端的外显肽,而目的蛋白作为 *N* 末端外显肽。表达的蛋白前体由于 Asp 残基被取代,缺少 Asp 环化和 *C* 末端剪切功能,在蛋白质内含肽的 *C* 末端不发生剪切作用,从而可以利用几丁质树脂的吸附加以纯化。改进后的表达载体是把目标蛋白融合在内含肽的末端,以 Glu 取代 *C* 末的 His 残基来调节内含肽 *C* 末端的剪接,从而快速纯化目的蛋白质。

3.2 直接检测蛋白质的表达及蛋白质之间的相互作用

Zhang 等^[17]用重叠的寡聚核苷酸构建了一个有效的微型内含肽(mini-intein)检测系统,在设计蛋白质内含肽的 *N* 末端和 *C* 末端之间插入亲和标记,把荧光蛋白(GFP)连在有亲和标记但缺乏自导引核酸内切酶区域的蛋白质内含肽的 *C* 末端,目的蛋白连在内含肽的 *N* 末端,这样利用荧光蛋白可以很容易的检测融合蛋白的表达。目前已有 17 种目的蛋白被检测,结果表明:目的蛋白的含量与总细胞的 GFP 蛋白荧光呈线性关系。另外,还可以利用这一系统检测蛋白质之间的相互作用^[18];把两个具有相互作用的目的蛋白分别融合进蛋白质内含肽的两个片段,再直接与分开的 EGFP (Enhanced green fluorescent protein)的 *C* 末端和 *N* 末端相连,蛋白剪接作用产生 EGFP 荧光团,这一方法简化了蛋白质相互作用的检测。

3.3 用于结构和功能的研究

X 衍射晶体学促进了内含肽结构的研究,但并不能解决催化机理的研究,也不能解决内含肽如何折叠及外显肽怎样影响折叠的,利用蛋白质的反式剪接现象可以解决这一问题。蛋白质的反式剪接(trans-splicing),是把两个分别表达和纯化的多肽链通过非共价连接融合进一个蛋白质内含肽的非功能区,组成具有蛋白剪接功能的内含肽。如:去除结核杆菌的 RecA 内含肽中间的自导引核酸内切酶区域,留下具有剪接功能的 440 个氨基酸残基,作为构建蛋白质剪接的元件,再通过生物化学和生物物理学的方法研究其结构和功能^[19]。研究发现,RecA 内含肽两侧的片段互补,用人工合成的 30~35 个氨基酸残基的肽段取代 RecA 内含肽的 *C* 末端相应肽段,可以促进蛋白质的反式剪接。

4 蛋白质剪接的研究进展

蛋白质剪接的发展及其研究进展吸引了细胞生物学、生物化学、分子生物学等不同领域学者的关注,并从不同的角度进行各种广泛的研究。近年来,在蛋白质化学合成的基础上,利用蛋白质剪接现象,发展了一种蛋白质融合方法——

表达蛋白连接(Expressed protein ligation,即 EPL^[19],利用结核杆菌的 RecA 的内含肽,将人工合成的肽链用化学方法连到内含肽的 C 末端。特别是在同一肽链中导入 N 末端的 Cys 残基和 C 末端的硫酯键,形成分子内连接反应,以合成环化蛋白质,增加蛋白质的稳定性。随着 EPL 的发展,在重组蛋白内含肽的两个末端均可以发生连接反应,这样可以将两个重组蛋白直接连在一起,还可以将合成的小肽链表达盒(cassette)特定地插入大分子蛋白中^[20]。

蛋白质剪接现象的发现,尽管只有短短的 10 多年时间,但它揭开了生物化学、生物物理学及蛋白质工程领域令人欣喜的新篇章。当前的工作极大地促进了我们对蛋白质剪接机理的了解,并可改变剪接点上、下游的序列,自主的设计蛋白质内含肽,调控蛋白质的剪接^[21];甚至用内含肽体外重组的多肽片段,也可以得到具有剪接功能和核酸内切酶活性的蛋白质内含肽元件^[22];但许多有关机理的详细过程及其作用还须做进一步研究,如:蛋白质剪接中的前 3 个反应速率比同类反应大得多,表明作为酸(碱)催化剂的残基可能促进这 3 个亲核反应;尽管许多蛋白质内含肽的保守序列已经测定,但它们在剪接中的具体作用不详。尽管利用蛋白质内含肽表达体系可以有效地断裂、纯化蛋白质,还需要选择合适的筛选标记,以分离、纯化蛋白质。在蛋白质工程中,利用蛋白质内含肽重组新的蛋白质时,要考虑新的蛋白质外显肽与内含肽的匹配及活性问题。另外,有关蛋白质剪接的酶学、细胞生理学及其进化方面的研究也少有报道。

参 考 文 献 (References):

[1] Hirata R, Ohsumi Y, 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*, 1990,265:6726~6733.

[2] Paulus H. Protein splicing; A novel form of gene expression and paradigm for self-catalyzed protein rearrangements. *Pure and Applied Chemistry*, 1998,70(1):1~8.

[3] Lambowitz A M, Belfort M. Introns as mobile genetic elements. *Annu Rev Biochem*, 1993,22:1125~1127.

[4] Cooper A A, Stevens T H. Protein splicing; self-splicing of genetically mobile elements at the protein level. *Trends Biochem Sci*, 1995,20(9):351~356.

[5] Fancine B, Perler FB. InBase: the Intein Database. *Nucleic Acids Research*, 2002,30(1):383~384.

[6] Liu X Q. Protein-splicing intein; Genetic mobility, origin and evolution. *Annual Review of Genetics*, 2000,34:61~76.

[7] Perler F B, Adam E. Protein splicing and its applications. *Curr Opin Biotechnol*, 2000,11(4):377~383.

[8] Xie J, Huang J F, Liu C Q. Analysis of the characteristic sequence of intein and revision of its motifs. *Chinese Sci Bull*, 2000,45(6):2525~2530.

[9] Dalgard J Z, Klar A J, Moser M J, Holley MD, Chatterjee A, Mian IS. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nuc Acids Res*, 1997,25(22):4626~4638.

[10] Degen Gimble F S. Degeneration of a homing endonuclease and its target sequence in a wild yeast strain. *Nucleic Acids Research*, 2001,29(20):4215~4223.

[11] Gimble F S, Thorner J. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, 1992,357:301~306.

[12] Lew B M, Mills K V, Paulus H. Protein splicing in vitro with a semisynthetic two-component minimal intein. *The Journal of Biological Chemistry*, 1998,273(26):15887~15890.

[13] Clarke N D. A proposed mechanism for the self-splicing of protein. *Proc Natl Acad Sci USA*, 1994,91(23):11084~11088.

[14] Vanderslice P, Copeland W C. and Robertus J D. Site-directed alteration of serine 82 causes nonproductive chain cleavage in prohistidine decarboxylase. *J Biol Chem*, 1988,263:10583~10586.

[15] Chong S, Montello G E, Zhang A, Cantor E J, Liao W, Xu M Q, Benner J. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucle Acids Research*, 1998,26(22):5109~5115.

[16] Zhang A H, Gonzalez S M, Cantor E J, Chong S R. Construction of a mini-intein fusion system to allow both direct monitoring of soluble protein expression and rapid purification of target proteins. *Gene*, 2001,275(2):241~252.

[17] Takeaki, Ozawa, Yoshio, Umezawa. Detection of protein-protein interactions in vivo based on protein splicing. *Current Opinion in Chemical Biology*, 2001,5(5):578~583.

[18] Lew BM, Paulus H, Mills KV. Characteristics of protein splicing in trans mediated by a semisynthetic split intein. *Biopolymers*, 1999,51(5):355~362.

[19] Xu MQ, Evans TC Jr. Intein-mediated ligation and cyclization of expressed proteins. *Methods*, 2001,24(3):257~277.

[20] Saburo Aimoto. Contemporary Methods for Peptide and protein synthesis. *Current Organic Chemistry*, 2001,5(1):45~87.

[21] Stu Borman. Controlled protein splicing. *Chemical & engineering news*, 2002,180(29):11.

[22] Amitai G, Pietrokovski S. Fine-tuning an engineered intein. *Nature biotechnology*, 1999,17(9):854~855.