

蛋白质内含子的特征、转移和演化

谢 君 刘次全* 黄京飞 石秀凡 邵 丹

(中国科学院昆明动物研究所, 昆明 650223)

摘要 自从 1990 年发现第一个蛋白质内含子, 就引起了有关专家的高度重视, 蛋白质内含子不仅在理论上丰富了遗传信息翻译后加工的内容, 而且实践上在蛋白质纯化方面有着广泛的应用前景. 鉴于蛋白质内含子的不断发现, 有必要对其在自我剪切, 传递, 演化等方面的进展做一概述.

关键词 蛋白质内含子, 自我剪切, 传递, 演化

学科分类号 Q617

蛋白质内含子是蛋白质中的一段多肽链, 靠自我剪切的方式从前体蛋白中分离出来, 同时两端的肽链以肽键的方式相连^[1], 蛋白质内含子中的外显子连接方式是不同于其他形式的蛋白质自我分解 (如糖基天冬氨酸酶或者 Hedgehog 蛋白的剪切)^[2,3]. 蛋白质内含子在古细菌、真细菌、真核生物中都有分布, 但从现在报道的蛋白质内含子来看, 古细菌中的蛋白质内含子最多, 真细菌中次之, 真菌中最少, 这一分布规律与 RNA 内含子相反. 另外其分布在物种上有选择性, 如真核生物中, 只发现在单细胞生物存在蛋白质内含子. 而且它们主要分布在与 DNA 复制, 转录, 翻译有关的蛋白酶或转移因子上. 大部分蛋白质内含子含有自导引核酸内切酶的特征序列, 有些还具有自导引核酸内切酶的活性. 序列比较^[4,5]和结构分析^[6~8]表明 N 端大约 100 个氨基酸和 C 端大约 50 个氨基酸残基对自我剪切是非常重要的. 这二个自我剪切区域由一段连接区序列 (大部分具有核酸内切酶特征) 分开. 本文从蛋白质内含子的序列特征、自我剪切的机制、转移、演化等方面的进展作一概述.

1 蛋白质内含子的特征序列

根据 <http://www.neb.com/inteins.html> 数据库, 蛋白质内含子有 8 个模体, 其中, A、B、F、G 属于自我剪切域的模体, 而 C、D、E、H 属于核酸内切酶域的模体. 少数蛋白质内含子不含核酸内切酶域, 或者含有的序列不具有核酸内切酶的特征. 但在含有核酸内切酶模体蛋白质内含子中, 核酸内切酶的序列比自我剪切域更为保守.

蛋白质内含子模体中, 没有绝对保守的残基,

除了模体 A 中 Cys/Thr/Ser, 模体 B 中 Thr^{**}Ser, 模体 G 中 His/Asn 相对比较保守, 模体中的其他位置一般为某一类性质的氨基酸残基, 特别是疏水氨基酸残基比较多. 模体 A 中的 Cys/Thr/Ser 的羟基在自我剪切的第一步转酯反应中, 将 N 端外显子接过来, 模体 B 中的 Thr-His 残基与主链原子形成氢键, 使蛋白质内含子前的肽键保持一种非标准的顺式构象, 从而辅助 N 端剪切位点的酰基转移^[8]. 蛋白质内含子模体 G 的倒数第二个残基 His 在 Asn 的环化和 C 端的切除反应中与 Asn 的羰基氧形成氢键, 导致这个肽键不稳定^[6,8,9]. 但最近报道, 蛋白质内含子也可以在蛋白质内含子的次末端残基不是 His 的条件下发生^[10], 在这种条件下, 与 Asn/Gln 羰基氧形成氢键可能是另外一个氨基酸残基, 因为这一残基只需要在三维结构上与 Asn/Gln 靠近, 而在一级序列上隔开很远, 也不会有多大的影响. 其他残基如果能形成相似的氢键, 也可以取代这些辅助残基而完成这些相同的功能. 以上蛋白质内含子的模体是在 1997 年已报道的 36 个蛋白质内含子序列搜索到的, 由于蛋白质内含子序列还在不断增多, 无疑会发现更多的模体.

2 蛋白质内含子的自我剪切过程

由于蛋白质内含子的剪切是一个快速的过程, 要了解自我剪切全过程是很困难的, 以下给出的自我剪切过程是根据蛋白质内含子 Mxe GyrA 和 Sce VMA 的晶体结构以及人工突变连接残基导致的自

* 通讯联系人.

Tel: 0871-5195183, E-mail: wudong11@263.net

收稿日期: 2000-05-11, 接受日期: 2000-07-26

我剪切过程中断,而产生一些剪切中间产物来推断的.现在一般认为,蛋白质内含子加上C端外显子的第一个氨基酸残基包含了足够的自我剪切信息,但外显子在静电荷和空间结构上对自我剪切的速度和效率产生影响.

总的来说,蛋白质自我剪切由3个在剪切处的保守残基的四步亲核攻击完成,还有一些酸性和碱性或氢原子结合残基辅助这四步亲核置换反应^[11, 12].

下面以两端连接残基是Ser残基为例来说明蛋白质内含子的自我剪切过程(图1).

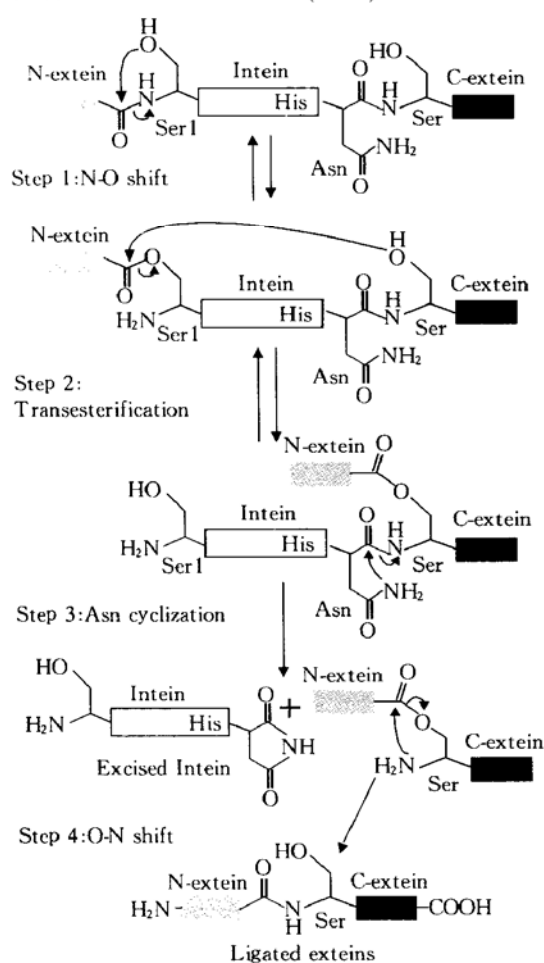


Fig. 1 The process of protein splicing

图1 蛋白质内含子自我剪切过程

第一步:由于蛋白质内含子N端的酰胺键,在模体B中的His Thr作用下形成不稳定的顺式构象,而与蛋白质内含子的N端Ser侧链上羟基发生反应导致蛋白质内含子的N端外显子转移到Ser的侧链上.已发现一些蛋白质内含子N端是Ala残基,既然Ala不含羟基或巯基的侧链,就不能象Ser/Thr/Cys一样能够完成酰基转移,但如果N端

连接肽键不稳定,而下游外显子第一个氨基酸残基在适当的位置可以攻击这个肽键,则可以直接将上游外显子转移到下游外显子的第一个氨基酸残基侧链上.

第二步,下游外显子的Ser/Thr/Cys攻击上游外显子的酯/硫酯键,导致上游外显子从蛋白质内含子的N端分离,而转移到下游外显子的Ser/Thr/Cys的侧链上,从而形成分支蛋白.

第三步,C端保守残基Asn环化形成琥珀酰亚胺环,导致C端剪切位点的断裂,琥珀酰亚胺能够水解形成Asn或者isoasparagine.有些蛋白质内含子C端是Gln,虽然这些蛋白质内含子还没有发现它们的自我剪切活性,但如Asn一样,Gln也能发生环化反应从而可以取代Asn而完成这一功能.

第四步,O-N或S-N酰基重组,在上下游外显子之间形成天然的肽键,这一过程是自发进行的.

3 蛋白质内含子和Hedgehog蛋白自我催化域的相似性

Hedgehog蛋白是胚胎形成过程中的信号分子^[13].它刚合成出来的时候,是没有活性的前体蛋白,它要经过与蛋白质内含子自我剪切相似的过程后才形成有活性的蛋白.这个前体蛋白由N端的信号区域和C端的自我催化区域(Hh-C)组成.

Hh-C第一个氨基酸残基是Cys,这个Cys起酰基转换的作用,将N端的信号域以酯键的方式转移到Cys的巯基上,这一反应过程与蛋白质内含子自我剪切过程的第一步相似.而且Hh-C也有与蛋白质内含子的模体A和B相对应的保守序列^[14].与蛋白质内含子的第二步的转酯反应相似,胆固醇的羟基攻击这个酯键,导致胆固醇以酯键的方式连接到Hedgehog蛋白的信号蛋白域上.晶体结构比较发现,Hh-C区域与蛋白质内含子的自我剪切区域非常相似,现在一般认为,Hedgehog自我催化域和蛋白质内含子是来源于同一祖先,是高等生物对蛋白质内含子连接两端肽键能力的修改.

4 自导引核酸内切酶和蛋白质内含子的转移

蛋白质内含子的转移是自导引核酸内切酶启动的.虽然自导引核酸内切酶与限制性核酸内切酶一样也是在基因的特定位置点上剪开DNA双链,但自导引核酸内切酶与一般的核酸内切酶在结构、识别特性,基因组上的位置都是不同的.

自导引核酸内切酶可分四个家族,它们的序列

特征分别为: LAGLIDADG, GIY-YIG, H-N-H 和 His-Cys. 自导引核酸内切酶在 DNA 上识别区域跨越 12~40 个碱基对, 而且相对于限制性核酸内切酶, 自导引核酸内切酶更能忍受在识别区域的碱基对的变化, 另外自导引核酸内切酶不象其他的核酸内切酶需要其他分子的辅助才能发挥作用, 最后, 自导引核酸内切酶在古细菌、真细菌、真核生物都有分布. 而限制性核酸内切酶只分布于古细菌、真细菌、某些真核生物的病毒中.

这种自导引核酸内切酶作用于蛋白质内含子基因的等位基因上, 在不含蛋白质内含子的等位基因上剪切一个缺口. 然后蛋白质内含子基因在重组酶、DNA 聚合酶、连接酶、分解酶的作用下^[15], 利用同源重组, 将蛋白质内含子基因的拷贝转移到不含蛋白质内含子的等位基因上. 这种转移方式跟 RNA I 类内含子的转移相似. 而且, 启动这种转移的都是自导引核酸内切酶.

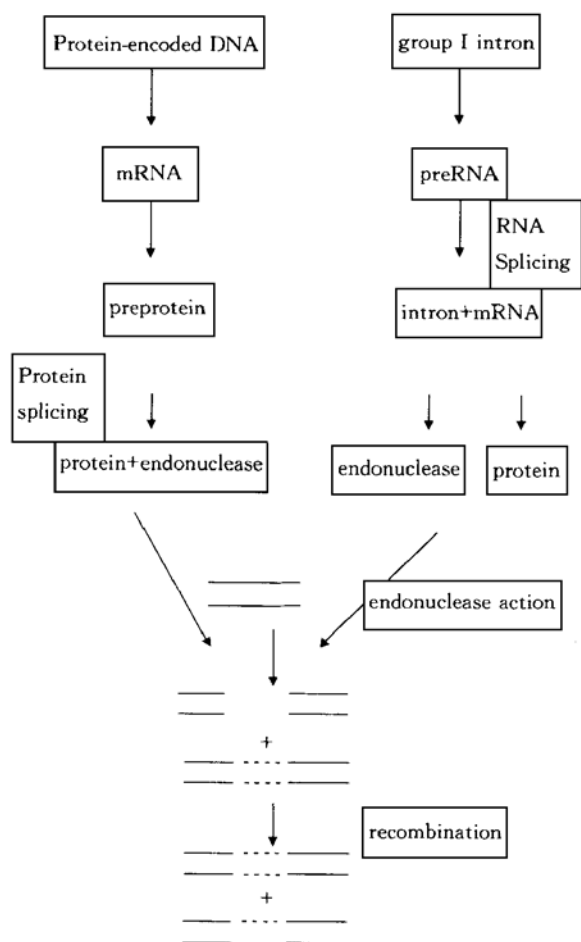


Fig. 2 The transfer of intein and intron

图 2 intein 和 intron 的转移

----: 自导引核酸内切酶编码区域; —: 蛋白质编码区域.

5 蛋白质内含子的演化

作者用 100 个蛋白质内含子的模体做同源比较和系统树分析表明, 绝大部分同源蛋白质内含子聚在一起, 这暗示这些蛋白质内含子来源于同一祖先, 但又看不出蛋白质内含子从古细菌到真细菌再到真核的演化过程, 而且有下列四个方面的证据更使得蛋白质内含子来源于同一祖先的学说无法解释: a. 有人发现蛋白质内含子的密码子不同于外显子的密码子^[16], b. 有些蛋白质内含子的相似性要远大于外显子的相似性, c. GB-D, GFJ thermococcales, *T. litoralis* 多聚酶相比, 虽然 GB-D 与 *T. litoralis* 多聚酶的相似性 (78%) 要小于 GB-D 和 GFJ thermococcales 分离株的 DNA 多聚酶的相似性 (98%), 但 GB-D 与 *T. litoralis* 多聚酶中含有蛋白质内含子 (蛋白质内含子的相似性达 60%), 而 GB-D 和 GFJ thermococcales 分离株的 DNA 多聚酶中却不含有蛋白质内含子^[17]. d. 在酵母中已经观察到蛋白质内含子的转移. 当不含蛋白质内含子的等位基因通过有性生殖、接合、转导、噬菌体感染、质粒转移等进入细胞后, 发现蛋白质内含子有转移的现象, 而且这种转移是在核酸内切酶剪切不含蛋白质内含子的等位基因, 然后通过同源重组而实现的^[18]. 这一系列现象都暗示蛋白质内含子是来源于不同物种而插入到外显子中的.

因此, 作者认同如下观点: 如果蛋白质内含子是有活性的自导引核酸内切酶, 那么蛋白质内含子更可能来源于转移. 如果是没有核酸内切酶活性的蛋白质内含子, 也可能是通过遗传得来的, 并随着时间的流逝可能丢失. 为了避免非活性突变, 这种丢失机制应该是非常特殊的. 同源重组可以导致蛋白质内含子的丢失, 但在单倍体生物中, 同源重组也只能在偶尔形成双倍体时才能发生.

参 考 文 献

- 1 Perler F B, Davis, E O, Dean, *et al.* Protein splicing elements: inteins and exons - a definition of terms and recommended nomenclature. *Nucleic Acids Res*, 1994, **22** (7): 1125~ 1127
- 2 Guan C, Cui T, Rao V, *et al.* Activation of glycosylasparaginase: Formation of active N-terminal threonine by intramolecular autoprolysis. *J Biol Chem*, 1996, **271** (3): 1732~ 1737
- 3 Porter J A, Ekker S C, Park W J. Hedgehog patterning activity: Role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell*, 1996, **86** (1): 21~ 34

- 4 Dalgaard J Z, Klar A J, Moser M J, *et al.* 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*, 1997B, **25** (22): 4626~ 4638
- 5 Pietrokovski S. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci*, 1998, **7** (1): 64~ 71
- 6 Duan X Q, Gimble F S, Quijcho F A. Crystal structure of PF-SceI, a homing endonuclease with protein splicing activity. *Cell*, 1997, **89** (4): 555~ 564
- 7 Hall T M T, Porter J A, Young K E, *et al.* Crystal structure of a hedgehog autoprocessing domains: Conservation of structure, sequence and cleavage mechanism between hedgehog and self-splicing proteins. *Cell*, 1997, **91** (1): 85~ 97
- 8 Klabunde T, Sharma S, Telenti A, *et al.* Crystal structure of gyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Struct Biol*, 1998, **5** (1): 31~ 36
- 9 Xu M, Perler F B. The mechanism of protein splicing and its modulation by mutation. *EMBO J*, 1996, **15** (19): 5146~ 5153
- 10 Chen L, Benner J, Perler F B. Protein splicing in the absence of an intein penultimate histidine. *J Biol Chem*, 2000, **275** (27): 20431~ 20435
- 11 Perler F B, Xu M Q, Paulus H. Protein splicing and autoproteolysis mechanisms. *Curr Opin Chem Biol*, 1997C, **1** (3): 292~ 299
- 12 Shao Y, Kent S B H. Protein splicing: occurrence, mechanisms and related phenomena. *Chem & Biol*, 1997, **4** (3): 187~ 194
- 13 Beachy P A, Cooper M K, Young K E, *et al.* Multiple Roles of Cholesterol in hedgehog protein biogenesis and signaling. *Cold Spring Harbor Symp Quant Biol*, 1997, **62**: 191~ 204
- 14 Koonin E V. A protein splice-junction motif in hedgehog family proteins. *Trends Biochem Sci*, 1995, **20** (4): 141~ 142
- 15 Belfort M, Perlman P S. Mechanisms of intron mobility. *J Biol Chem*, 1995, **270** (51): 30237~ 30240
- 16 Liu X Q, Hu Z M. Identification and characterization of a cyanobacterial DnaX intein. *FEBS Lett*, 1997, **408** (3): 311~ 314
- 17 Perler F B, Kumar S, Kong H. Thermostable DNA polymerases. In: Adams M W W, ed. *Enzymes and Proteins from Hyperthermophilic Microorganisms*. New York: Academic Press, 1996. 377~ 435
- 18 Gimble F S, Thorner J. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, 1992, **357** (6376): 301~ 306

Recent Advance in the Study of Intein

XIE Jun, LIU Ci-Quan*, HUANG Jing-Fei, SHI Xiu-Fan SHAO Dan

(Kunming Institute of Zoology, The Chinese Academy of Science, Kunming 650223, China)

Abstract Since the first intein was found, more and more attention were paid on it. It not only enrichs the content of the process that the gene transfers its information but also can be used in protein purification. The recent advance in the sequence characteristic, transfer, evolution and the mechanism of splicing of intein was summarized.

Key words intein, self-splicing, transfer, evolution

* Corresponding author. Tel: 86-871-5195183, E-mail: wudong11@263.net

Received: May 11, 2000 Accepted: July 26, 2000