

蛋白质剪切及其应用

宋丽萍 黄华琚

(中国科学院遗传与发育生物学研究所, 北京 100101)

摘要 蛋白质剪切是一种翻译后修饰事件, 它将插入前体蛋白的中间的蛋白质肽段(Intein, internal protein fragment)剪切出来, 并用正常肽键将两侧蛋白质多肽链(Extein, flanking protein fragments)连接起来。在此过程中不需要辅酶或辅助因子的作用, 仅需四步分子内反应。Intein 及其侧翼序列可以通过突变产生高度特异性的自我切割用于蛋白质纯化、蛋白质连接和蛋白质环化反应, 在蛋白质工程方面有广泛的应用前景。

关键词 蛋白质内含子, 蛋白质剪切, 环状肽, 纯化, 蛋白质工程

中图分类号 Q617 文献标识码 A 文章编号 1000-3061(2003)02-0249-06

20 世纪 90 年代, 蛋白质剪切的发现开辟了生物化学的新篇章。自从第一个蛋白质剪切元件——芽殖酵母中的 VMA intein 基因被报道以来, 在真细菌、古细菌、单细胞真核生物中已经陆续发现了 118 种 intein^[1]。蛋白质剪切元件——intein 对于蛋白质工程来说是一个功能强大的工具。Intein 载体不仅可用于纯化、表达毒性蛋白, 而且可以产生 IPL (Intein mediated protein ligation) 所需的活性末端, 引入在核糖体生物合成过程中不能加入的非天然的氨基酸(如磷酸化或糖基化修饰)、标签、发色基团、部分修饰或标记的蛋白质, 研究结构—活性之间的关系。此外, intein 还可以有效的介导蛋白质环化, 提高蛋白质的稳定性。本文试从 intein 的结构、蛋白质剪切机制、intein 在蛋白质纯化、蛋白质连接、蛋白质环化等方面的应用作一概述。

1 Intein 的结构及蛋白质剪切的机制

Intein 是一段存在于前体蛋白的蛋白序列, 是在蛋白质剪切过程中被剪切出来的片段。大部分 intein 均具有双功能: 蛋白质剪切活性和自导引核酸内切酶(Homing endonuclease)活性。自导引核酸内切酶活性是位点特异性的, 在识别序列和功能上与核酸内含子的自导引核酸内切酶类似, 可以在无 intein 的等位基因的双链 DNA 处定点切割, 引起 intein 的插入^[2]。

See VMAI intein 的 X-射线晶体衍射图谱显示它具有双结构域。结构 I 包含 intein 序列的 N 末端(前 182 个氨基酸)和 C 末端(最后 44 个氨基酸), 几乎均为 β 折叠, 为剪切结构域。结构 II 位于二者之间, 为自导引核酸内切酶结构域^[3]。

突变研究及序列数理统计研究表明, intein 普遍拥有这种双结构域结构^[4,5]。See VMAI intein 以及 *Mtu* RecA intein 的自导引核酸内切酶结构域的缺失可产生仅具有蛋白质剪切功能的微 intein^[6,7]。*Mxe* GyrA 在天然状态下就缺乏内切酶结构域, 但仍具有蛋白质剪切活性。这些发现提示 intein 的 C 端及 N 端可组装成蛋白质剪切的功能性结构域, 而位于中间的结构域对于蛋白质剪切并不重要。但也有例外, 在 *Psp* Pol-I intein 内, 核酸内切酶的结构域不同大小的缺失均可以导致失去剪切活性^[8]。

Intein 有 10 个保守的 intein 模体(图 1)^[9]。其中 4 个位于自导引核酸内切酶结构域中。Intein 和下游 extein 的第一个残基提供了蛋白质剪切的几乎所有的必要条件。对 intein 的多重序列比较发现了 4 个在剪切邻近区域保守的残基。(i) Intein N-末端的半胱氨酸, 苏氨酸或丙氨酸。(ii) Intein C-末端的天冬酰胺。(iii) 位于下游 extein 的 N 末端第一个残基处的半胱氨酸、苏氨酸或丝氨酸。此外, intein 倒数第二个组氨酸可辅助 intein C 末端的切割反应^[10]。

蛋白质剪切有 3 个明显的特点(1) 蛋白质剪切完全由蛋白质内含子的氨基酸介导。(2) 蛋白质剪切是一个分子内过程。(3) 蛋白质剪切不需要辅酶或代谢能, 其过程主要涉及键的重排。

蛋白质剪切分为 4 步(图 2)(1) 在剪切体的上游, intein 的第一个残基半胱氨酸或丝氨酸(Ser1)的侧链发生 N-S 或 N-O 酰基重排, 产生具有易断裂肽键的线性酯键中间体, 在中性 pH 条件下, 酰胺键-酯键之间的平衡趋向于形成酰胺键。这种硫酯键关联在巯基试剂存在的条件下, 易发生亲核

收稿日期 2002-10-17, 修回日期 2002-12-20。

基金项目 国家高科技研究发展计划项目资助(No. 863-102-09-44-01)

* 通讯作者。Tel: 86-10-64857285; Fax: 86-10-64857285; E-mail: hlhuang@genetics.ac.cn

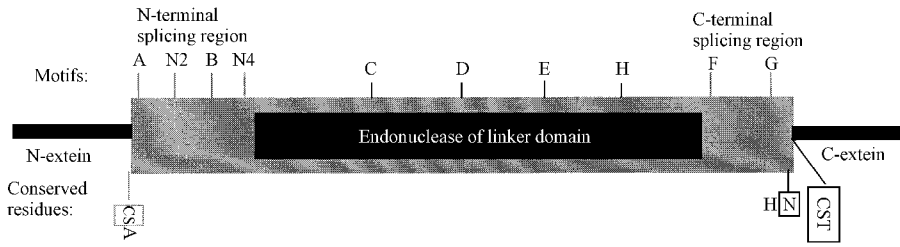


图 1 蛋白质内含子模体^[9]

Fig. 1 InteIn motif^[9]

The N-terminal and C-terminal splicing region form the splicing domain. Homing Endonuclease separate the splicing regions.

Blocks A, N2, B, N4, F, G are the splicing motifs. Nucleophiles are boxed^[9]

置换, 形成具有硫酯键末端的靶蛋白。如果替换 intein N 末端的半胱氨酸或丝氨酸 (Ser1), N 端剪切位肽键的切割就会完全被封阻。(2) 通过剪切位点 C 末端的半胱氨酸或丝氨酸攻击线性酯键中间体形成分枝酯键中间体。剪切连接处下游半胱氨酸的巯基或丝氨酸, 苏氨酸的羟基亲核攻击上游剪切处的前面的羰基碳原子, 导致 N 端 extein 转移到 C 端 extein 第一个残基的侧链上。这种分支结构的形成是可逆的。(3) InteIn C 末端剪切处的天冬酰胺环化形成琥珀酰亚胺环, 同时分枝酯键中间体断裂, 产生带有 C 末端琥珀酰亚胺环的 intein 和通过酯键连接的两个 extein。倒数第二个氨基酸组氨酸对于邻近天冬酰胺的环化有辅助作用。组氨酸替换成其它氨基酸时, 会影响 C 末端的切割反应。(4) 琥珀酰亚胺环水解, extein 之间的酯键发生 O-N 或 S-N 酰基重排, 产生稳定自然的肽键^[11]。

2 InteIn 介导的蛋白质纯化

Intein 有很多通用的模体, 提示它们有共同的进化起源^[9]。然而, 每个 intein 在其各自的宿主中都经过很长时间的进化来优化剪切过程, 因此对各自剪切位点邻近的氨基酸有一定的喜好。对于这些在 intein 中参与肽键形成、断裂的残基的研究为 intein 的应用打下了坚实的基础。Intein 与 extein 几乎没有序列同源性, 天然 extein 常常可以替换成外源的蛋白质而不影响 intein 的剪切活性。这一点是 intein 作为融合蛋白进行蛋白质纯化的理论基础。我们可以将靶基因与 N 或 C 末端修饰过的 intein 融合, 纯化标签嵌在相反方向或者 intein 的内部, 融合蛋白表达、纯化后, 通过诱导 intein 的切割活性, 释放出靶蛋白(图 3)。基于 intein 的蛋白纯化技术与其它系统相比, 该系统无需使用蛋白酶, 避免了蛋白酶切割所带来的非特异性酶解, 也无需进一步纯化去除蛋白酶, 提供了一个省时经济的方法^[12]。根据蛋白质剪切位点氨基酸替换的类型, 蛋白质剪切可以分为 intein N 末端剪切和 intein C 末端剪切(图 3)。到目前为止, 商业化的表达载体中利用的 intein 有 *Mxe* GyrA, *Mth* RIR1, *See* VMA。下面就两种纯化策略举例说明。

2.1 在靶蛋白的 C 末端融合 intein, 在 intein N 末端剪切(图 3 左)

pTYB^[12]载体将靶蛋白融合在 *See* VMA intein 的 N 端, 几丁质结合蛋白纯化标签 intein 融合在 intein 的 C 端。*See* VMA intein C 末端剪切位点的天冬酰胺被突变成丙氨酸, 封阻 intein C 末端的自剪切, 仅允许由半胱氨酸 N-S 酰基转移介导的 intein N 末端的切割。亲核试剂 DTT、2-巯基乙醇或半胱氨酸可以通过攻击硫酯键, 诱导 intein N 末端的切割。将靶蛋白融合在连有几丁质结合蛋白纯化标签的 intein 的 N 末端, 通过几丁质纯化融合蛋白前体后, 在还原剂存在的条件下诱导切割获得目的蛋白^[12]。

靶蛋白的产量取决于在细胞抽提物中获取的融合蛋白的产量及其柱上结合后的切割效率。在大肠杆菌中表达异源蛋白, 尤其是当靶蛋白为真核生物来源的时候, 常常会产生包涵体或者由于密码子使用的偏好不同而导致不表达。这些问题在 N 末端融合外源蛋白时就更加突出。低温诱导可以减少包涵体的形成, 提高融合蛋白的折叠及可溶性。切

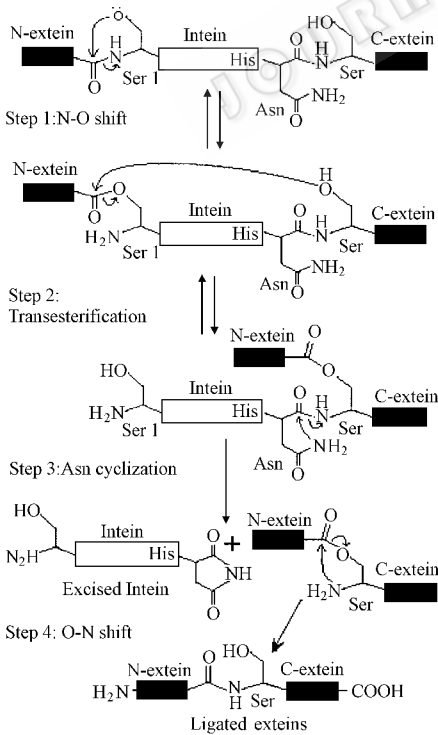
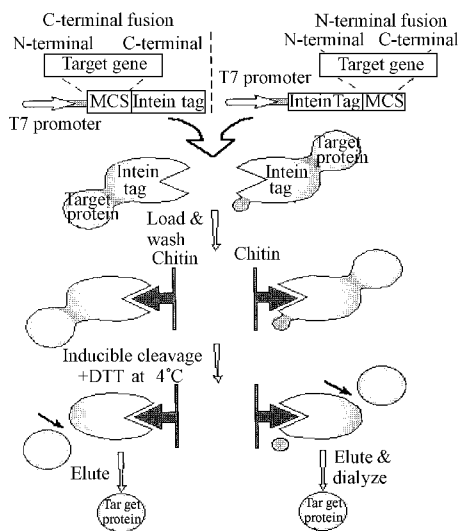


图 2 蛋白质剪切的机制^[9]

Fig. 2 The mechanism of protein splicing^[9]

图3 Intein 介导的蛋白质纯化示意图^[11]Fig. 3 Intein mediated protein purification^[11]

割效率主要受与 intein 切割位点邻近的靶蛋白的氨基酸序列的影响。以麦芽糖结合蛋白为例,20 多种氨基酸中的绝大多数氨基酸都不会影响融合蛋白的纯化及 DTT 介导的切割。当靶蛋白 C 末端的氨基酸为天冬氨酸、谷氨酸、精氨酸、组氨酸和苏氨酸时,体内切割效率提高,会导致融合蛋白的产量下降。当靶蛋白 C 末端的氨基酸为亮氨酸、异亮氨酸、缬氨酸时,只有在高温和长时间的诱导才能诱导融合蛋白的切割。如果 intein 附近为脯氨酸、半胱氨酸或天冬酰胺时,DTT 诱导的切割根本无法进行^[12]。因此,在使用这种载体时,往往需要向剪切位点附近添加额外的甘氨酸,降低体内的切割效率,提高体外的切割效率。

2.2 在靶蛋白 N 末端融合 intein,在 intein C 末端剪切(图 3 右)

pTYB11^[12]载体将 Sce VMA intein 倒数第二个组氨酸突变为谷氨酰胺,削弱了 intein C 末端的天冬酰胺介导的切割反应,同时将 C 端蛋白质 extein 的第一个氨基酸半胱氨酸突变为丙氨酸,封闭体内剪切。亲和标签插入在 intein 自导引核酸内切酶结构域处,不影响 intein 剪切结构域的结构和剪切活性。此外,在 intein 的 N-末端还融合了一小段 10 个氨基酸左右的 N 端 extein 用来提供翻译起始位点来优化表达。靶蛋白的 N 末端紧邻 intein 的 C 末端。这种表达载体表达的融合蛋白前体产量较高,体内不发生剪切反应,剪切位点的切割完全在体外来进行。

靶蛋白 N 末端的氨基酸会影响 intein C 末端的剪切^[13]。绝大多数氨基酸在 16℃、40h 或者 23℃、16h,可以介导 > 50% 的切割,在 4℃ 时,蛋氨酸、丙氨酸、谷氨酰胺可以介导有效的切割,而缬氨酸、异亮氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸、组氨酸会抑制 intein C 末端的切割。靶蛋白 N 末端的色氨酸、半胱氨酸和苏氨酸可以导致体内剪切,使融合蛋白的产量下降^[12]。

3 Intein 介导的蛋白质连接 (IPL) 和表达蛋白连接 (EPL)

引入非天然的氨基酸来改变蛋白质的结构及功能,将会大大提高我们对于蛋白质的理解,为生物医学研究提供新的工具,创造出新的治疗药物。例如蛋白质的位点特异性磷酸化修饰、引入荧光标签、自旋共振探针、交联试剂等可帮助我们深入地了解生物学过程,然而目前引入非天然氨基酸的方法仍然困难重重^[14]。在天然化学连接中,一个含有 N 末端的半胱氨酸的肽与一个拥有 C 末端硫酯键的蛋白质连接,在连接位点形成天然肽键^[15]。然而,到目前为止,用于天然化学连接的含有 C 末端硫酯键的肽仅能通过化学合成来产生,而该方法在技术上要求苛刻,且有片段大小的限制 (< 15kD)。

蛋白质剪切可以产生重组的 C 末端硫酯键。蛋白质剪切的第一步 N-S 或 N-O 酰基转移, N 端 extein 转酯至位于 intein 氨基端的半胱氨酸/色氨酸/苏氨酸残基的 SH 或 OH 侧链上。基于这个原理,我们可以设计突变,使 intein 仅仅能促进蛋白质的剪切的第一步^[16]。融合表达在这样的 intein 氨基末端的蛋白可以通过分子内硫酯键的转酯反应被巯基试剂 (MESNA, thiophenol) 切割,产生重组的 α 硫酯键^[17]。含有氨基端半胱氨酸残基的蛋白和拥有羧基端 α 硫酯键的肽在含有巯基试剂的条件下可以产生硫酯键介导的化学浓缩,然后 S-N 酰基转移,形成天然的肽键^[18]。这个过程叫 intein 介导的蛋白质连接 (IPL)。

用于 IPL 的蛋白片段既可以是化学合成的也可以是基因表达产物。最初, IPL 仅用于将拥有羧基端 α 硫酯键的重组蛋白与合成的含有氨基端半胱氨酸残基的肽段相连接,形成荧光标记的蛋白、特异性修饰的蛋白(如磷酸化),合成细胞毒性蛋白质 (RNaseA, Hpa I 内切核酸酶)^[19, 20]。Muller^[14] 用 EPL 方法产生含有磷酸化的酪氨酸肽段来研究磷酸化对于 src 酪氨酸激酶活性和构象的影响。

IPL/EPL 的下一步飞跃是在天然条件下连接两个表达蛋白^[21]即表达蛋白连接 (EPL, Expressed protein ligation)(图 4),运用该技术区段性用同位素标记蛋白,大大提高了 NMR 蛋白质结构的分辨率。同时,连接多个合成或表达的肽段也成为可能。以 C 末端片段起始,顺序地向其 N 端加入片段,每个片段必须拥有一个羧基端的 α 硫酯键和一个氨基端的被屏蔽的半胱氨酸。当连接反应发生后,这个半胱氨酸通过蛋白酶消化屏蔽而进行下一轮的连接。

基于 intein 的蛋白质纯化载体通过产生具有反应活性末端的蛋白质扩展了天然化学连接的应用范围。在核糖体生物合成过程中不能加入的如磷酸化或糖基化的非天然的氨基酸、标签、发色基团等等都可以使用 IPL 技术引入到蛋白质中去。IPL/EPL 加速了对于腺体-配体结合、转录后修饰对于酶活稳定性、亚单位相互作用的生化及生物物理分析的研究进程。

4 Intein 介导的蛋白质环化

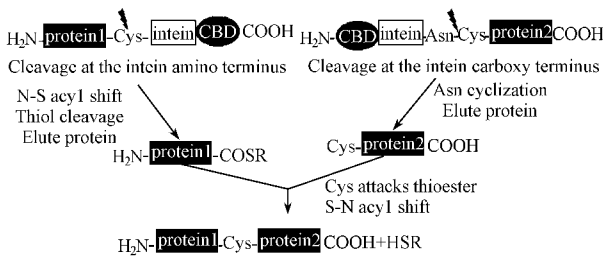


图4 表达蛋白连接

Fig.4 Expressed protein ligation

的青睐。(1)稳定性好。这是由于环化的蛋白质可减少非折叠状态的构象熵值。(2)折叠速率快。这是由于其减少了折叠途径的数目。(3)对于氨基端和羧基端的蛋白酶具有抗性^[22]。环化的蛋白因其不同寻常的热稳定性和生物活性在医药和工业上备受重视。

Golderberg and Creighton 通过化学交联方法,形成肽骨架 N 末端和 C 末端的肽键而产生了第一个环化蛋白-牛胰腺胰岛抑制肽(BPTI)^[23],然而这种方法并不适用于结构并不十分清楚的大蛋白质。最近,intein 已经被用来介导蛋白质或肽的环化。可以分为两种方式进行环化。

4.1 体外环化

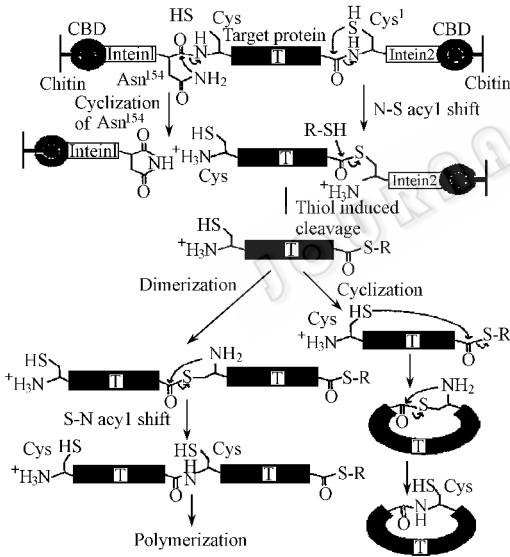


图5 Intein 介导的体外环化^[24]

Fig. 5 In vitro cyclization pTWIN System^[24]

The TWIN system sandwiches a target protein (T) between two inteins and can produce an N-terminal cysteine and a C-terminal thioester on the same bacterially expressed protein. Purification of the precursor protein is simplified by using a CBD, which binds to chitin resin. Cleavage at the C terminus of Intein 1, releases an N-terminal cysteine on the target protein. A thiol reagent induces cleavage of Intein 2 and releases the target protein with an activated C-terminal thioester. Two reaction pathways can occur, either an intermolecular reaction that leads to polymerization or an intramolecular reaction to generate a circular protein. In both possibilities a peptide bond is formed during the ligation reaction

其中一个方法便是在同一个表达蛋白中产生具有 N-intein-靶蛋白-C intein 形式的融合蛋白。在体外经巯基试剂及 pH 诱导切割,产生 N 末端的半胱氨酸和 C 末端的硫酯键,然后 N 末端和 C 末端反应,产生环化蛋白或串联蛋白^[24]。基于这一原理的 pTWIN^[24]系统(图 5)已经成功地环化了 E. coli 硫氧还蛋白(12kD)、麦芽糖结合蛋白(42kD)、脑结合肽(9aa)、血小板凝集抑制剂 RGD(10aa)、CDR-H3/C2(16aa)^[11]。另一种方法使用 intein N 末端的切割载体来表达靶蛋白。将位点特异性的蛋白酶识别序列融合在靶蛋白 N 末端的半胱氨酸的邻位,表达纯化出融合蛋白后,在体外经蛋白酶切产生 N 末端带有半胱氨酸的靶蛋白-intein-CBD 的融合蛋白,然后 N 末端半胱氨酸攻击 intein C 末端硫酯键,形成环化分子^[11]。这种自环化的反应的成功与否在很大程度上都取决于活性位点的暴露情况和靶蛋白的折叠状态如何。

4.2 体内环化

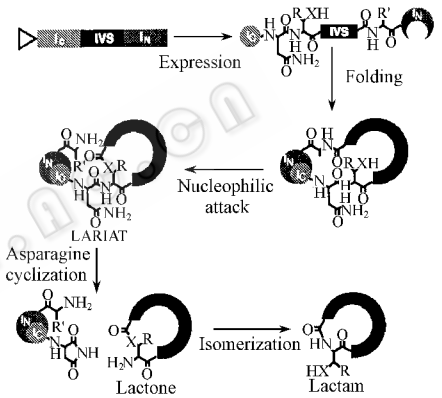


图6 Intein 介导的体内环化^[26]

Fig. 6 Intein mediated in vivo cyclization (SICLOPPS)^[26]

A chimeric gene is created that interposes an intervening sequence (IVS, target protein) between genes for the C-intein (IC) and N-intein (IN) components of an intein. Expression in a suitable host yields a fusion protein that folds to generate an active intein. The intein catalyzes the attack of a side chain nucleophile from the carboxy-terminus of the IVS upon the carbonyl carbon of the amino acid at the amino-terminus of the IVS to generate a lactone lariat intermediate. The cyclic product is liberated as a lactone by cyclization of the asparagine residue at the carboxy-terminus of IC

蛋白质和肽的体内环化对于在体内大规模产生环化的蛋白质和肽至关重要。这个技术同时也为在体内产生大规模的用于功能筛选和生物物理鉴定的环化肽库提供了可能。因为它可以在生物体内产生更稳定的没有副产物的环化蛋白,便于体内功能筛选和工业化生产^[25]。体外环化蛋白因其需要通过在体外化学处理产生 C-末端的硫酯键中间体而受到限制。这种困境可以通过利用人工或者是天然存在的分裂的 intein 的蛋白质的反式剪切活性得以解决。基因重建试验说明 intein 可以缺失自导引核酸内切酶结构域而不影响 extein 的剪切活性,intein 可以分开表达为两个多肽链,通过非共价结合而重新获得剪切的活性。通过将靶蛋白的基因表达在分裂内含子的中间,表达具有 C 端 intein,待环化的目

标蛋白-N端 intein 结构的融合基因,在体内通过分子内的反式剪切作用,靶蛋白的氨基端和羧基端在体内通过正常的肽键连接而产生环化产物(图6)。这种方法叫做体内分裂 intein 介导的肽或蛋白质的环化(SICLOPPS)^[12]。理论上讲,这种策略不仅适用于天然分裂蛋白,也适用于具有相似结构的人工改造蛋白。到目前为止报道的 intein 中,仅有3个适于在体内产生高水平的环化靶蛋白: *Synexohystis sp.* PCC6803. DnaE intein^[26], *Pyrococcus furiosus*. PI-PfuI intein^[26], *M. tuberculosis* 的 recA intein^[22], 其中 *Synexohystis sp.* PCC6803. DnaE intein 由于其两端的肽段具有固有的亲和力,这种天然的 intein 显然更适于介导体内环化^[27]。作者所在实验室正致力于用 *Synexohystis sp.* PCC6803. DnaE intein 来介导重组单抗的体内环化。我们用分裂的 Ssp DnaE intein 将靶蛋白夹在中间,在大肠杆菌中表达,期望获得稳定的环化蛋白质。目前我们已经成功地证明了其在体内介导的剪切作用,但环化仍有待于进一步证明。

5 结 论

Intein 是蛋白质肽链中插入的序列。作为蛋白质水平而不是 RNA 水平剪切的结果,它们在核酸水平上存在,在成熟的基因产物中消失。自从在 20 世纪 90 年代被发现,在 intein 的化学及结构基础方面以及 intein 的应用等诸多方面已经取得了令人瞩目的成就。对 intein 的蛋白质剪切及切割活性的调控为我们提出了蛋白质表达和操作的全新策略。改造的 intein 已经用来产生独特的反应基团而广泛地应用于表达蛋白纯化、连接、环化等方面。利用 intein 区段性标记蛋白扩展了 NMR 结构分析对于蛋白质大小的限制,并为应用 NMR 分析特异区域的构象改变及大分子量的多体蛋白质提供了可能^[27,28]。Intein 介导的环化提高了蛋白质的稳定性及生物活性。Intein 载体产生的蛋白质 C 末端的硫酯键在蛋白质化学中具有重要的价值。尽管 intein 的应用尚属起步阶段,但从近几年发表的文章来看, intein 已经迅速地成为一个多功能的工具。我们拭目以待它在蛋白质工程方面作出更大的贡献。

REFERENCES(参考文献)

[1] <http://bioinformatics.weizmann.ac.il/~pietro/inteins/>

[2] Perler F B. Protein splicing of inteins and hedgehog autoproteolysis: structure, function and evolution. *Cell*, 1998, **92**(1): 1-4

[3] Duan X, Gimble F S, Quijcho F A. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*, 1997, **89**: 555-564

[4] Kawasaki M, Nogami S, Satow Y *et al.* Identification of three core regions essential for protein splicing of the yeast Vma1 proteozyme. A random mutagenesis study of the entire VMA1-derived endonuclease sequence. *J Biol Chem*, 1997, **272**(25): 15668-15674

[5] Nogami S, Satow Y, Ohya Y, Anraku Y. Probing novel elements for protein splicing in the yeast VMA1 proteozyme: a study of replacement mutagenesis and intragenic suppression. *Genetics*, 1997, **147**(1): 73-85

[6] Chong S, Xu M Q. Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs. *J Biol Chem*, 1997, **272**

(25): 15587-15590

[7] Derbyshire V, Wood D W, Wu W *et al.* 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**(21): 11466-11471

[8] Wu H, Xu M Q, Liu X Q. Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim Biophys Acta*, 1998, **1387**(1-2): 422-432

[9] <http://www.neb.com/neb/inteins.html>

[10] Southworth M W, Adam E, Panne D *et al.* Control of protein splicing by intien fragment reassembly. *The EMBO Journal*, 1998, **17**(4): 918-926

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

[12] Xu M Q, Paulus H. and Chong S. Fusions to self-splicing inteins for protein purification. *Methods Enzymol*, 2000, **326**: 376-418

[13] Chong S, Montello G E, Zhang A *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**(22): 5109-5115

[14] Muir T W, Sondhi D, Cole P A. Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A*, 1998, **95**(12): 6705-6710

[15] Dawson P E, Muir T W, Clark-Lewis I, Kent S B. Synthesis of proteins by native chemical ligation. *Science*, 1994, **266**(5186): 776-779

[16] Cotton G J, Muir T W. Peptide ligation and its application to protein engineering. *Chem Biol*, 1999, **6**(9): R247-256

[17] Chong S, Mersha F B, Comb D G *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, 1997, **192**(2): 271-281

[18] Noren C J, Wang J, Perler F B. Dissecting the chemistry of protein splicing and its applications. *Angew Chem Int Ed Engl*, 2000, **39**(3): 450-466

[19] Evans T C Jr, Benner J, Xu M Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci*, 1998, **7**(11): 2256-2264

[20] Severinov K, Muir T W. Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J Biol Chem*, 1998, **273**(26): 16205-16209

[21] Xu R, Ayers B, Cowburn D, Muir T W. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc Natl Acad Sci U S A*, 1999, **96**(2): 388-393

[22] Siebold C, Erni B. Intein-mediated cyclization of a soluble and a membrane protein *in vivo*: function and stability. *Biophys Chem*, 2002, **96**(2-3): 163-171

[23] Goldenberg D P, Creighton T E. Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J Mol Biol*, 1983, **165**(2): 407-413

[24] Evans T C Jr, Benner J, Xu M Q. The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J Biol Chem*, 1999, **274**(26): 18359-18363

[25] Iwai H, Lingel A, Pluckthun A. Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-PfuI intein from *Pyrococcus furiosus*. *J Biol Chem*, 2001, **276**(19): 16548-16554

[26] Scott C P, Abel-Santos E, Wall M, Wahn D C, Benkovic S J. Production of cyclic peptides and proteins *in vivo*. *Proc Natl Acad Sci*

- [27] Otomo T , Teruya K , Uegaki K , Yamazaki T , Kyogoku Y . Improved segmental isotope labeling of proteins and application to a larger protein . *J Biomol NMR* , 1999 , **14** (2) : 105 - 114
- [28] Otomo T , Ito N , Kyogoku Y , Yamazaki T . NMR observation of se-

lected segments in a larger protein : central-segment isotope labeling through intein-mediated ligation . *Biochemistry* , 1999 , **38** (49) : 16040 - 16044

Protein Splicing and Its Application

SONG Li-Ping HUANG Hua-Liang*

(*Institute of Genetics and Development Biology , Chinese Academy of Sciences , Beijing 100101 , China*)

Abstract Protein splicing is a newly discovered posttranslational editing process that removes an internal protein fragment from the protein precursor. During the splicing process the internal protein fragment , intein , triggered the self-excision from the precursor protein and the concomitant ligation of the flanking protein fragments , exteins . The self-catalysis requires neither auxiliary enzymes nor cofactors and only involves four intramolecular reactions. A number of key catalytic residues in inteins and flanking fragments have been identified , which led to the development of the protein splicing process as a protein engineering tool. Controllable cleavage of the peptide bond at either the N or the C terminus of an intein has allowed the design of novel strategies for manipulation of protein and peptides. Affinity purification of recombinant proteins can be facilitated by fusion the target protein with an intein. The fusion also creates C-terminal thioester , which expands the scope of chemical ligation in protein. Inteins can be engineered in a “ split and inverted ” configuration to form a cyclic polypeptide consisting of the sequence linking two intein subdomains. This article summarizes the recent advance in the mechanism of protein splicing and its applications in protein purification , protein ligation and protein cyclization.

Key words protein splicing , intein , protein purification , cyclic peptide , protein ligation

Received : 10-17-2002

This work was supported by Grant from the State “ 863 ” High Technology R&D Project of China (No. 863-102-09-44-01) .

* Corresponding author. Tel 86-10-64857285 ; Fax 86-10-64857285 ; E-mail : hlHuang@genetics. ac. cn

© 中国科学院微生物研究所期刊联合编辑部 <http://journals. im. ac. cn>