

# 信号肽及其在蛋白质表达中的应用

韦雪芳 王冬梅 刘思 周鹏

(热带作物生物技术国家重点实验室, 华南热带农业大学海口校区, 海口 571101)

**摘要:** 分子生物学研究已进入后基因组时代, 其中心任务是更多地关注基因组表达的蛋白质的结构和功能。由于基因功能最终通过其表达产物——蛋白质来实现, 因此, 要了解基因组全部功能活动, 最终也必须回到蛋白质上。另外, 在菌株、培养和发酵等逐渐成熟的条件下, 构建高效的表达载体以提高外源蛋白质的表达量是降低工业化生产成本的关键。随着研究的深入, 发现信号肽对蛋白质的定位有着非常重要的作用, 使得信号肽的研究不仅具有重要的理论意义, 而且也具有潜在的应用价值。就信号肽的结构和功能, 信号肽的捕获方法及其在原核表达系统和真核表达系统中表达外源蛋白质的应用做一些介绍。

**关键词:** 信号肽 捕获 原核表达系统 真核表达系统 蛋白质表达

## Signal Sequence and Its Application to Protein Expression

Wei Xuefang Wang Dongmei Liu Shi Zhou Peng

(National Key Biotechnology Laboratory for Tropical crops, South China University of Tropical Agricultural, Haikou 571101)

**Abstract:** The study of molecule biology is coming to terminal genome times now. The important aim is to pay attention to the structures and functions of proteins. The functions of gene were implemented by expression productions——proteins, so if want to know the functions and activities of genome, we must study proteins first. On the other side, when strains, culture and ferment conditions are mature, then make up of effective expression vectors is the key of reduce the cost of industrial manufacture. Peptides play an important role in the orientation of proteins in pace of further study. Because of these, the study of peptides is not only important in theory, but also worth in applications. Signal peptides structures and functions, signal sequence traps and the proteins expressions in prokaryote and eukaryote expression systems were illustrated in the article.

**Key words:** Signal peptide Signal sequence trap Prokaryote expression system Eukaryote expression system Protein expression

蛋白质分子中的信号肽研究始于 20 世纪 60 年代, 是引导新合成肽链转移到内质网上的一段多肽, 位于新合成肽链的 N 端, 一般由 15~30 个氨基酸残基组成, 含有 6~15 个带正电荷的非极性氨基酸, 由于信号肽又是引导肽链进入内质网腔的一段序列, 又称开始转移序列(start transfer sequence)。外源蛋白在宿主菌, 如大肠杆菌中的表达形式多为细胞内不溶性表达(包涵体), 少数为细胞外分泌表达。利用信号肽来引导外源蛋白定位分泌到细胞特定区间, 提高可溶性, 可避免因包涵体复性带来的困难。目前研究采用的信号肽来自表达系统自身的信号序列或外源信号

序列, 或两者兼而有之。研究表明, 多种外源基因连接上信号肽后, 在原核表达系统, 如大肠杆菌、L 型细菌、芽孢杆菌和乳酸杆菌中等都得到了分泌表达; 信号肽也广泛应用于真核表达系统如毕赤酵母和昆虫杆状病毒表达系统中<sup>[1]</sup>。

### 1 信号肽的结构与功能

#### 1.1 信号肽的结构

穿过合成所在的细胞到其他组织细胞去的蛋白质, 统称为分泌性蛋白质(secretory proteins)。信号肽位于分泌蛋白的 N 端, 一般由 15~30 个氨基酸组成

收稿日期: 2006-06-08

作者简介: 韦雪芳(1981-), 女, 广西人, 华南热带农业大学硕士研究生, 微生物专业

通讯作者: 周鹏(1963-), 男, 研究员, 主要从事植物和生物技术研究, Tel: 0898-66988564, E-mail: zhp6301@126.com

(表 1), 包括三个区: 一个带正电的 N 末端, 称为碱性氨基末端; 一个中间疏水序列, 以中性氨基酸为主, 能够形成一段 螺旋结构, 它是信号肽的主要功能区; 一个较长的带负电荷的 C 末端, 含小分子氨基酸, 是信号序列切割位点, 也称加工区。当信号肽序列合成后, 被信号识别颗粒(SRP)所识别, 蛋白质合成暂停或减缓, 信号识别颗粒将核糖体携带至内质网上, 蛋白质合成重新开始。在信号肽的引导下, 新合成的蛋白

质进入内质网腔, 而信号肽序列则在信号肽酶的作用下被切除<sup>[2]</sup>。近年来一些研究表明, 许多分泌蛋白的移位信息虽确由一部分疏水肽段所携带, 但这一部份肽段可以不在 N 端, 如终止转运序列存在于新生肽链的 C 端, 也可以不被信号肽酶切除, 如卵清蛋白含有内部信号肽, 它的前体与成熟形式都没有被信号肽酶切除的过程, 其 N-端氨基酸结构在第 9 位有带电基团, 疏水结构并不明显。

表 1 三种蛋白质的信号肽序列

蛋白质	信号序列
Preproalbumin	Met-Lys-Trp-Val-Thr-Phe-Leu-Leu-Leu-Phe-Ile-Ser-Gly-Ser-Ala-Phe-Ser Arg...
PreIgG light chain	Met-Asp-Met-Arg-Ala-Pro-Ala-Gln-Ile-Phe-Gly-Phe-Leu-Leu-Leu-Phe-Pro-Gly-Thr-Arg-Cys Asp...
Prelysozyme	Met-Arg-Ser-Leu-Leu-Ile-Leu-Val-Leu-Cys-Phe-Leu-Pro-Leu-Ala-Ala-Leu-Gly Lys...

1.2 信号肽的功能

在布洛贝尔“信号假说”的启示和有关蛋白质氨基酸序列的比较研究基础上, 一些新的信号肽先后被发现。这些信号肽负责不同类型蛋白质的新生肽链的定位。信号肽的功能, 已不仅决定一个蛋白质是否为分泌蛋白, 而且和蛋白质或其新生肽链在细胞内的全方位的定位有关。

新生肽链或蛋白质中, 一些残基的化学修饰也是转译后加工的一个重要内容。发生修饰的残基决不是任意的, 也和肽链中的氨基酸序列密切相关。从这个意义而言, 这些和残基修饰有关的肽段, 也可认为是残基修饰的信号肽。例如, 在肽链中连接有 N-糖链的天冬酰胺残基(N), 一定是位于 Nx(S/T)这种特定三肽序列中的天冬酰胺; 一些蛋白质的 C 末端附近可以为萜类所修饰, 能接上萜类的半胱氨酸酸残基, 同样是具有特定序列的 C 末端四肽中的半胱氨酸酸残基, 而且其中某些残基还决定了所接上的萜类的长度。为此, 可认为在肽链中尚存在着与残基修饰相关的信号肽。

一些肽段的存在与含该肽段肽链的降解有关, 这类肽段可以视为肽链降解的信号肽。如细胞质中某些带有 KFERQ/RIDKQ 序列的蛋白质易于进入溶酶体, 然后在那里被降解。又如, 细胞质内快速被降解的蛋白质通常含有 PEST 四肽序列<sup>[3]</sup>。

2 信号肽的捕获方法

分泌性蛋白在生物个体发育、生理功能的发挥及各种病理过程中起着重要作用。而分泌性蛋白的分泌有赖于蛋白质 N 端的信号肽的存在。信号肽捕获系统(signal sequence trap, SST)就是依据分泌性蛋白质的这一特性进行基因克隆的功能性基因克隆系统, 由 Tashiro K 等于 1993 年建立, 该方法可以快速、大量地克隆真核细胞编码分泌性蛋白的基因序列, 目前用该系统已经获得了大量编码分泌性蛋白的基因<sup>[4]</sup>。

SST 法包括两个基本步骤: (1) 构建并转染和报告基因融合的 cDNA 文库; (2) 筛选、分离携带胞外标记分子的细胞。胞外标记分子的出现意味着 cDNA 片段编码功能性信号肽。SST 法具有的优势是一个普通功能性框架包含了众多的分泌性和 I 型膜蛋白, 既指导初级多肽链进入分泌通道的 N 末端信号肽, 同时省去了烦琐的生物功能分析过程。不足之处是由于需要利用流式细胞仪和复杂的免疫染色技术筛选细胞, 大大地限制了该方法的使用和效率。

其后, 学者们在标记分子选择、表达细胞以及信号序列来源等多方面进行了改进。Klein<sup>[5]</sup>等利用酵母在蔗糖培养基中生长需要转化酶的原理, 建立了一种利用酵母作为转化细胞的替代方法。该方法虽然简便, 但有些真核细胞表达的信号不能被酵母系统识别。Kojima<sup>[6]</sup>等根据凝血酶原蛋白受体(MLP)的性质,

建立了通过逆转录病毒介导表达筛选的信号序列捕获系统(SST-REX),该系统能高效捕捉信号肽序列。Chen<sup>[7]</sup>设计了一种以人碱性磷酸酶(PLAP)基因为标记分子的 PST 系统,只有编码起始和融合在 PLAP 框内的功能信号肽才能把嵌合蛋白运送到细胞表面或分泌入胞浆,从而利用底物-酶结合的特性进行比色分析。Moffatt<sup>[8]</sup>利用病毒颗粒建立了 EB-VSS 法,在这一方法中,标记分子的表达可以触发目标核酸运出细胞到培养基中,从而避免了筛选细胞的必要。Péerly 等<sup>[9]</sup>利用基因组 DNA 代替 cDNA 创立了外显子信号捕获(signal-exon trap, SET)法,该法适用于几乎所有的表达类型和表达强度的分泌性和膜性信号捕获。Gebauer 等<sup>[10]</sup>利用融合在大肠杆菌新霉素磷酸转移酶表达框内的 CD2 表达抗原作为标记分子基因,建立了一种筛选编码分泌性和膜性蛋白调节基因整合分子的逆转录病毒介导的信号序列捕获策略,利用这一策略可以有效地从基因组水平筛选编码分泌性和膜性蛋白整合分子<sup>[11]</sup>。

### 3 在蛋白质表达中的应用

#### 3.1 在原核表达系统中的应用

大肠杆菌表达系统因其生长周期短,操作简单,遗传学、生理学特性明确等优点而应用广泛,但在表达外源蛋白时,常形成无活性的包涵体,需要通过复性等复杂步骤才可能恢复部分活性。外源蛋白的分泌表达不仅有利于表达蛋白的纯化、减小毒蛋白对细胞的毒害,更重要的是蛋白在分泌的过程中能获得正确折叠和二硫键化,从而获得有活性的蛋白质。

大肠杆菌的分泌型蛋白定位于内膜、外膜、周质空间和胞外环境,它们在 N 端或 C 端带有一定的结构(包含着分泌信号),其中 N 端带有信号肽分子的蛋白,其在跨越内膜时得到 Sec 家族蛋白因子协助,在跨膜过程中信号肽可能被切除<sup>[12]</sup>。C 端带有分泌信号的蛋白,如 HlyA(-溶血素),其信号序列位于 C 端最后的 46-50 个残基内,其重组融合蛋白不具任何毒性,且在转运后信号肽不被切除。这一信号序列已成功用于 IL-1、二氢叶酸还原酶等几十种蛋白的正确分泌<sup>[13]</sup>。

目前采用的信号肽一般来自大肠杆菌的外膜蛋白(如外膜蛋白 A OmpA、OmpF、噬菌体受体 LamB、热稳定肠毒素 ST 等)或周质蛋白(如碱性磷

酸酶 PhoA、麦芽糖结合蛋白 MBP、DsbA 等),此外金黄色葡萄球菌蛋白 A 和胡萝卜软腐欧文氏果胶酶裂解酶(PelB)的信号肽也常被采用。

枯草芽孢杆菌是传统的工业生产菌,能分泌大量的胞外蛋白,是研究蛋白质表达和分泌的良好系统,利用它进行工业规模的外源基因表达和分泌,产生有用的蛋白类产品,有很重要的经济价值。枯草芽孢杆菌的信号肽序列长度大多在 18 到 35 个氨基酸之间,并且不具有一致序列。其特定的典型结构特征是具有带正电的 NH<sub>2</sub> 端,随后是一段疏水区域(H 区)和一个极性更强的 C 区,C 区带有一致的酶切识别位点 Ala-X-Ala,酶切发生于羧基端丙氨酸之后。两个丙氨酸偶尔可以被其他具有短侧链的氨基酸代替,X 位偏爱具有巨大侧链的氨基酸。枯草芽孢杆菌信号肽中的疏水区对于蛋白质的分泌有着重要的影响,适当的疏水性有利于蛋白的分泌,疏水区过长或过短都不利于蛋白的分泌。有研究表明,信号肽疏水性能提高促进青霉素 G 酰化酶分泌<sup>[14]</sup>。信号肽的选择有时能明显地影响分泌表达水平,如枯草芽孢杆菌蛋白酶、中性蛋白酶、崖边杆菌 RNA 酶(bamase)和果聚糖酶等 4 种酶的信号肽,以芽孢杆菌 BE1510 为宿主菌,分别表达重组链霉抗生物素蛋白(streptavidin),结果中性蛋白酶信号肽的分泌效率最高。

虽同为原核生物,但大肠杆菌和芽孢杆菌的信号肽在蛋白质定位上还有所差别。有报道将 pGPB14 双功能载体插入相同的启动子-信号肽序列后,分别转化大肠杆菌和枯草芽孢杆菌。结果表明,在大肠杆菌中克隆的启动子-信号肽序列在大肠杆菌和枯草芽孢杆菌中虽都具有启动基因表达和蛋白质分泌的功能,但在大肠杆菌中,-N 酰胺酶主要积累在周质空间,细胞内及细胞外的酶活力都很低;而枯草芽孢杆菌中,酶活力主要分泌到胞外,在周质空间和细胞内积累很少,这主要是两者细胞壁结构差异所致<sup>[1]</sup>。

#### 3.2 在真核表达系统中的应用

真核生物不但有细胞核、细胞质和细胞膜,而且还有许多膜性结构的细胞器,像原核细胞一样,真核细胞合成的蛋白质 N 端也有信号肽,其也能形成两个螺旋的发夹结构,这个结构可插入到内质网膜,正在合成的多肽链带和内质网内腔内。

真核表达系统中应用较为广泛的为酵母表达系



统和昆虫杆状病毒表达载体系统,而酵母表达系统中研究较清楚、应用最广的是 *Pichia.Pastoris*。目前酵母表达系统中常使用的信号肽包括某些外源蛋白自身的天然信号肽、-因子信号肽(-MF)、蔗糖酶基因 SUC2 信号肽序列、酸性磷酸酶基因 PHO1 的信号肽序列、间质金属蛋白酶(MMP)/间质金属蛋白酶组织抑制剂信号肽序列<sup>[15]</sup>。其中 -因子信号肽(-MF)使用最广。-因子信号序列由 87 个氨基酸组成,是来自 *S.cerevisiae* 的性成熟因子前导序列,并且已将这段序列编码的信号肽插入到几个 *P.Pastoris* 的表达载体中(Invitrogen),用于外源基因的分泌表达,如 pGAPZ -A、B、C 与 pMET -A、B、C 等。在构建 -因子信号肽融合基因时,需保留 KeX<sub>2</sub> 蛋白酶切位点附近的谷氨酸-丙氨酸(Glu-Ala)间隔区, Glu-Ala 的存在会避免错误切割的发生<sup>[16]</sup>。

信号肽序列的调整对外源蛋白质在毕赤酵母中的表达有一定的影响。如信号肽使用酵母偏爱密码子能提高植酸酶基因的表达式<sup>[17]</sup>。在酿酒酵母交配因子(MF)信号肽中间插入 EEAEAEAEPK 共 10 个氨基酸的信号肽序列能使胰岛素的分泌率提高 2 倍以上<sup>[18]</sup>。在酿酒酵母的因子信号肽序列 MF4I 的 N 端分别引入 1~10 个毕赤酵母 Aox1 蛋白质的 N 端氨基酸,构成 10 种不同的分泌信号肽序列,用于植酸酶基因的分泌表达,结果以 N 端加入 A、I、P 三个氨基酸的信号肽序列提高的分泌表达量最大<sup>[19]</sup>。

酵母细胞对信号肽序列的识别有一定的灵活性,在一定程度上可以识别外源蛋白的信号肽进行蛋白输送,但利用外源蛋白的天然信号肽引导表达的重组蛋白虽可分泌,但效率较低,因此,在一定程度上,需要利用酵母本身的蛋白信号肽来引导外源蛋白的表达<sup>[20]</sup>。如人血清蛋白(HSA)和牛凝乳酶(Rennin)用自身的信号肽序列即可在 *Pichia.Pastoris* 中成功表达<sup>[21]</sup>。然而,对于一些蛋白的自身信号肽 *Pichia.Pastoris* 不能有效利用,如反转录酶<sup>[22]</sup>。

除此之外,酵母表达系统产物的另一个问题是信号肽切除不完全,原因是细胞内合成大量外源蛋白质时,酵母二氨基肽酶的活力不足以除去它们氨基末段的 4 个附加氨基酸残基,酵母菌自我平衡调节机制受到破坏,出现信号肽切除不完全现象<sup>[23]</sup>。研究表明,通过启动子和目的基因连接处寡核苷酸定向缺失诱变

导致含有 N 天然端成熟目的蛋白释放的方法,可以解决信号肽加工不完全的问题<sup>[24]</sup>。

昆虫杆状病毒表达载体系统自 1983 年建立以来,已经发展成为较完善的真核表达系统,成功的表达了多种外源基因。但信号肽在昆虫表达系统中的应用还很有限。

目前在昆虫表达载体中对信号肽在基因工程应用上的研究结果主要有: 1) 信号肽能被信号肽酶正确识别和切割,能促进产物的分泌,如利用 -干扰素信号肽和 IL-3 信号肽构建的昆虫杆状病毒分泌载体,含蜂蜜毒肽的信号肽载体能提高木瓜蛋白酶的分泌效率 5~9 倍<sup>[25]</sup>; egt 和 gp67 信号肽与 HIV-1 的 gp120 基因融合在昆虫表达载体中可以提高产物分泌效率 6~20 倍。2) 信号肽能使产物分泌至细胞外,但在不正确位点切割,如利用 MAG 信号肽表达的睫状体神经营养因子(CNTF)<sup>[26]</sup>。3) 信号肽不能引导分泌表达,如淀粉酶的信号肽不能引导木瓜蛋白酶的分泌表达<sup>[25]</sup>。4) 信号肽不能被切除,如用 gp67 信号肽引导慈姑蛋白酶抑制剂(API)基因在昆虫表达载体中表达虽获得了高效表达,且表达产物的 90%以上能分泌到细胞外,但信号肽在表达过程中都没有被切除<sup>[27]</sup>。

总之,信号肽在外源蛋白质的表达中具有重要作用,信号肽的提出为蛋白质的研究开辟了新的领域:蛋白质的空间属性。这为下一步研究蛋白质的结构和功能奠定了一定的基础。到目前为止,信号肽打破了最初“信号假说”赋予其的概念和作用——位于 C 端的信号肽先后被发现;其作用还与蛋白质或其新生肽链在细胞内的全方位的定位有关。这使得信号肽在生产药用蛋白质中倍受关注,但外源蛋白的纯化除了需要该蛋白高效的分泌外,更重要的因素之一还是选用适宜的信号肽。所以在实际研究中,应根据目的基因的不同,表达载体的不同而选用不同的信号肽,以达到实验或研究的目的。

#### 参考文献

- 1 郑斌,詹希美. 生物技术通讯, 2005, 16(3): 296~298.
- 2 罗进贤,李文清,等. 遗传学报, 1994, 21(1): 74~80.
- 3 王克夷. 生命的化学, 1995, 15(3): 1~4.
- 4 Tashiro K, Nakamura T, Honjo T. Methods Enzymol, 1999, 303: 479.
- 5 Klein RD, Gu Q, Goddard A, et al. Proc Natl Acad Sci USA,

- 1996, 93: 7108~7113.
- 6 Kojima T, Kitamura T. *Nat Biotechnol*, 1999, 17: 487~490.
- 7 Chen H, Leder P. *Nucleic Acids Res*, 1999, 27: 1219~1222.
- 8 Moffatt P, Salais P, Gaumond MH, et al. *Nucleic Acids Research*, 2002, 30( 19): 4285~4294.
- 9 Péterfy N, Gyuris T, Takács L. *Nucleic Acids Research*, 2000, 28( 7): 26~27.
- 10 Gebauer M, Melchner HV, Thomas B. *Genome Research*, 2001, 11: 1871~1877.
- 11 张奉学, 余传信. *Journal of Tropical Medicine*, 2004, 4( 4): 498~500.
- 12 袁宇, 甘人宝. *生物化学与生物物理进展*, 1995, 22( 4): 295~300.
- 13 龙建银, 王会信. *生物化学与生物物理进展*, 1997, 24( 2): 126~132.
- 14 杨运桂, 徐京宁, 等. *生物化学与生物物理学报*, 2000, 32( 2): 163~168.
- 15 Sreekrishna K, Brankamp RG, Kropp KE, et al. *Gene*, 1997, 190( 1): 55~62.
- 16 韩雪清, 刘湘涛, 等. *微生物学杂志*, 2003, 23( 4): 35~40.
- 17 Bei JL, Chen Z, Yang L, Liao L, Wang XZ, Jiang ZY. *Chinese J Biotechnol*, 2001, 17( 3): 254~258.
- 18 Kjeldsen T, Pettersson AF, Hach M. *Biotechnol Appl Biochem*, 1999, 24( 1): 79~86.
- 19 熊爱生, 彭日荷, 等. *生物化学与生物物理学报*, 2003, 35( 2): 154~160.
- 20 赵慧, 郑文岭, 等. *生命的化学*, 2003, 23( 3): 177~179.
- 21 范乔, 刘宏迪. *微生物学通报*, 1999, ( 4): 304~306.
- 22 Tschopp JF, Sverlow G, Kosson R, et al. *BioTechnology*, 1987, 5: 1305~1308.
- 23 刘文. *生物技术通讯*, 1998, 9( 1): 68.
- 24 Singh A, Lugovoy JM, Kohr WJ, et al. *Nucl Acid Res*, 1984, 12( 3): 8927.
- 25 Tessier DC, Thomas DY, Khouri HE, et al. *Gene*, 1991, 98: 177~83.
- 26 Murphy CL, McIntire JR, Davis DR, et al. *Protein Expr Purif*, 1993, 4: 349~357.
- 27 季平, 查新民, 等. *江苏大学学报*, 2003, 23( 2): 10~16.
- 28 魏虹, 王金富. *江西饲料*, 2004, ( 4): 8~11.
- 29 孙强, 王翼姝, 等. *遗传学报*, 2001, 28( 4): 397~384.
- 30 叶方寅. *国外医学分子生物学分册*, 1999, 21( 6): 377.
- 31 赵明, 张衡. *生理科学进展*, 1998, 29( 3): 226~230.
- 32 钱伟. *生物学教学*, 2001, 26( 3): 7~8.
- 33 欧阳立明, 张惠展, 等. *生物化学与生物物理进展*, 2000, 27( 2): 151~154.
- 34 李键仔. *生物学通报*, 2005, 40( 3): 21~23.

## 欢迎订阅 2007 年《生物技术通报》

《生物技术通报》(双月刊)创刊于 1985 年,是我国生物技术领域的国家级综合性科技期刊和中国农业核心期刊以及中国核心期刊遴选数据库来源期刊、中国学术期刊综合评价数据库来源期刊、中国科学引文数据库来源期刊等。

主要报道国内外农、林、牧、渔及医学等领域中生物技术研究的新进展、新技术、新成果与发展趋势,内容包括基因工程、细胞工程、酶工程、发酵工程、蛋白质工程以及生物工程的应用、新的实验技术与方法等。本刊内容丰富详实。欢迎广大读者到当地邮局订阅或直接与编辑部联系订阅(免邮资费),我们将按期及时邮寄。

本刊为双月刊、大 16 开本、80 页,每期定价 15 元,全年 90 元,邮发代号为 18-92,国外发行代号为 BM5607。

联系地址: 100081 北京中关村南大街 12 号信息所《生物技术通报》编辑部

电 话: 010-68919925(可传真)

E-mail: biotech@mail.caas.net.cn