

综述

外源基因在酵母中表达产物的分泌

李育阳

(复旦大学遗传学研究所, 上海)

近年来研究酵母表达外源基因的报道较多^[1-24]。作为基因工程中的表达系统, 分泌外源蛋白的能力是一个十分重要的特性, 因为产物的分泌避免了产物在细胞内的大量积累, 而产物的大量积累常常会影响细胞的正常生长及产物的稳定性。其次, 产物的分泌有利于产物的分离提纯, 这不仅因为提取产物不必破碎细胞, 还因为培养液的组分主要是小分子物质, 酵母在正常情况下分泌蛋白质的种类不多, 量也较少, 这样产物相对于杂蛋白的比例也就比较高。此外, 许多基因工程的产物, 特别是那些和医药有关的产品, 原来多数是分泌蛋白。最初合成的蛋白质只有经过伴随蛋白质的分泌过程进行的翻译后加工(包括切去信号肽、糖基化及空间折叠等)才能成为有生物活性的物质。因此蛋白质的分泌过程同产物活性有无或大小也是至关重要的^[15,20,22]。

(一) 酵母细胞分泌蛋白质的途径

目前已知的酵母分泌蛋白质有十几种^[19]。其中一部分可以直接分泌到培养基。如: 酵母的外激素, α 因子, a 因子和杀死毒素(Killer toxin)等。另一部分则只分泌到细胞的周质空间, 如: 酸性磷酸酯酶、蔗糖酶等。Schekman 和 Novick^[19] 从酵母选得一系列温度敏感的分泌缺陷突变体。在分析了各种突变体的蛋白质分泌性质后, 推测酵母细胞分泌蛋白质的途径是: 内质网-高尔基体-囊泡-胞外。即: 在翻

译形成蛋白质的同时借助于 N 端的信号肽转移到内质网, 在内质网内加上 N 连接的核心寡糖(N-linked core oligosaccharide)。在高尔基体样的复合物内进一步作糖基化修饰。然后再被包装进分泌囊泡。最后送至细胞外。信号肽在分泌蛋白质转入内质网时被专一性蛋白酶切去。

在蛋白质分泌过程中信号肽起着重要的作用。对已知结构的八种酵母分泌蛋白质的信号肽所作的分析表明: 酵母的信号肽和其他原核、真核生物的信号肽相似, 具有以下特点: (1) 高度的疏水性; (2) 长度在 17 到 30 个氨基酸残基之间; (3) 在信号肽中至少有一个带正电荷的氨基酸残基, 但位置并不一定在信号肽的 N 端最初 5 个氨基酸残基中。

(二) 外源前体蛋白质基因在酵母中的分泌表达

酵母细胞在一定程度上可以识别外源分泌蛋白质的信号肽进行蛋白质的输送, 分泌表达产物。表 1 是文献报道的一些结果。但是在多数情况下, 这种用编码前体蛋白质(Preprotein)的基因进行表达的水平一般比较低, 产物的分泌也较少。例如在 α 干扰素的情况下, 如果采用带信号序列的前体蛋白质基因进行表达, 只有不足 10% 的产物在培养基内, 真正分泌的干扰素只占总蛋白的 0.01%。用前体蛋白质

本文于 1986 年 6 月 5 日收到。

表 1 外源基因在天然信号肽指导下的分泌表达

表达基因	表达产物 分子量	表达信号		表 达 量	分泌量	加工情况	参考文献
		启动子	终止子				
细菌 β -内酰胺酶		β -内酰胺酶基因	β -内酰胺酶基因	是细菌中表达量的1/500	$<0.2\%$ (总蛋白)	加工所得的成熟蛋白和在细菌中相同。	16,17(1981)
人 α_1 干扰素	19.5kd	PGK	2 μ	2.7×10^7 u/L	4 %	细胞内及周质空间中正确加工的产物占34%，培养基中正确加工的产物占64%	9 (1983)
人 α_2 干扰素	19.5kd	PGK	2 μ	$1.9-2.8 \times 10^7$ u/L	3—8 %	培养基中正确加工的产物占45%	9 (1983)
人 γ 干扰素	17kd	PGK	2 μ	$0.9-1.9 \times 10^6$ u/L	10—21%		9 (1983)
小牛凝乳酶原	40.7kd	PGK	PGK	总蛋白的5 %		只有25%左右的产物酸化后被激活	13 (1983)
小麦 α -淀粉酶	42kd	PGK	α -淀粉酶基因		30—60%		18 (1984)
人 α_1 -抗胰蛋白酶	44kd	ARG3	ARG3	总可溶蛋白的0.3%	无分泌	多数产物不被加工。有糖基化，但程度不齐	4 (1984)
植物甜味蛋白 thaumatin	22kd	GAPDH	GAPDH			能正确加工prepro顺序	6 (1985)
小鼠免疫球蛋白 λ	~28kd	PGK	PGK	500ng/ml (可溶蛋白)	10—40%	在同时产生 λ 和 μ 的细胞中可以检测到抗体活性	24 (1985)
小鼠免疫球蛋白 μ	63.5kd	PGK	PGK	15ng/ml (可溶蛋白)	5—15%		
小牛凝乳酶原	40.7kd	GAL1		2.5mg/g (可溶蛋白)	$<6\mu$ g/g		22 (1985)
牛生长激素		GAL1		3.0mg/g (可溶蛋白)	3 μ g/g		22 (1985)

基因进行表达，水平低的原因可能是由于前体的折叠(Folding)不正确，影响了翻译效率或提高了对蛋白酶的敏感性^[18]。

在信号肽的加工方面，一般不能做到完全正确。在干扰素的情况下，切割位置正确的分泌蛋白质占了64%，其余的分泌蛋白质则多了三个从信号肽来的氨基酸，表达产物中的一部分虽然未能完全实现分泌，仍留在细胞内，但也进行了部分加工。还有些前体蛋白质基因在表达时，酵母细胞完全不能识别切割位点，蛋白质也未能分泌^[4,6]。总之，利用外源分泌蛋白质信号肽，通常只能得到加工过的，未经加工的以及加工不正确的三类蛋白质的混合物。

(三) 外源基因在酵母分泌信号肽指

导下的分泌表达

为了要取得对外源基因表达产物较好的分泌效果，通常需要用酵母自己的信号肽来指导。目前被采用的有 α -因子，SUC2和PHO5基因产物的信号肽。表2所列有关这方面工作的一些概况。

1982年Kurjan和Herskowitz首先克隆了编码酵母外激素 α -因子的基因MF α_1 ^[10]，随后有人就构建了以MF α_1 的信号序列来指导的分泌型表达载体。这个基因所编码的蛋白质有22个氨基酸长的信号肽，另有61个氨基酸的原片段(Pro-segment)，它的功能还不太清楚，其上有三个糖基化位点，可能它们与前体经加工直接进入培养基而不停留在周质空间有关。

表 2 外源基因在酵母信号肽指导下的分泌表达

表达基因	表达产物 分子量	表达信号		表达量	分泌信号	分泌量	加工情况	参考文献
		启动子	终止子					
β -内啡肽 (合成)	31个氨基酸	α -因子	α -因子		α -因子	150—200 $\mu\text{g/L}$	加工识别点在lys arg之后, *9,*19 lys前有断裂点	2 (1984)
Con 1 α -干扰素 (合成)	19.5kd	α -因子	α -因子		α -因子	2×10^8 u/L	加工识别点在lys arg之后, 有小于5%的产物降解	2 (1984)
人表皮生长因子	6 kd	α -因子	α -因子	细胞可溶蛋白的 7 %	α -因子	>95%	加工识别点在lys arg之后	3 (1984)
细菌 β -半乳糖苷酶	121—171kd	SUC2			SUC2	只分泌至内质网	只能进行N连接的核心寡糖基化	7 (1984)
α_1 干扰素	19.5kd	α -因子	2 μ 质粒	1×10^8 u/L	α -因子	>80%	加工在 lys arg 之后	21(1984)
鼠白细胞间介素-2	17kd	α -因子	Trp1	>10 $\mu\text{g/L}$	α -因子	>80%分泌至周质空间和培养基	加工识别点在 lys arg之后	12(1985)
凝乳酶原	40.7kd	GAL1		3mg/g (可溶蛋白)	SUC2	27 $\mu\text{g/g}$ (可溶蛋白)	在SUC2 部分有糖基化	22(1985)
凝乳酶原	40.7kd	SUC2		0.5mg/g (可溶蛋白)	SUC2	21 $\mu\text{g/g}$ (可溶蛋白)		
凝乳酶原	40.7kd	TPI		2.5mg/g (可溶蛋白)	SUC2	23 $\mu\text{g/g}$ (可溶蛋白)		
凝乳酶原	40.7kd	α -因子		2.0mg/g (可溶蛋白)	α -因子	11 $\mu\text{g/g}$ (可溶蛋白)		
凝乳酶原	40.7kd	GAL1		3.0mg/g (可溶蛋白)	PHO5	<6 $\mu\text{g/g}$ (可溶蛋白)		
凝乳酶原 (整合一个拷贝)	40.7kd	TPI		0.7mg/g (可溶蛋白)	SUC2	82 $\mu\text{g/g}$ (可溶蛋白)		
凝乳酶原 (整合四个拷贝)	40.7kd	TPI		3.0mg/g (可溶蛋白)	SUC2	240 $\mu\text{g/g}$ (可溶蛋白)		

在酵母基因组中, α -因子多肽编码区重复四次, 每一重复之间有编码间隔肽 (Spacer peptide) 的序列, 其中有 lys AAA arg AGG, 接下去是编码 2—3 个二肽 glu ala或asp ala的序列。形成 α -因子多肽的加工过程包括三个步骤: (1) 用 Cathepin B 或胰蛋白酶样的酶在 lys arg 后切割。(2) N-末端的二肽用二肽氨肽酶 (Dipeptidylaminopeptidase) 切除。(3) C-末端的碱性残基用羧肽酶 B (Carboxypeptidase B) 去除。

利用 α -因子的启动子和分泌信号已经表达和分泌了多种外源基因, 尽管表达水平相差很大, 但是一般都能得到比较好的分泌。分析表达产物的结构, 蛋白质加工过程的切割识别点和对 α -因子的加工过程一样。但是看来由二肽氨肽酶作用的步骤是一个限制加工速度的因素, 如果把编码成熟蛋白质的序列连接于 lys arg AAA AGG 之后, 这样加工过程就不需要二肽氨肽酶的作用了, 得到的就是结构正确的成熟蛋白质产物, 对分泌产物的糖基化过程, 还研究得不多, 一般通过分泌途径可以实现 N-

连接的核心寡糖的糖基化。

利用 SUC2 和 PHO5 基因的信号肽来指导外源基因的分泌工作还不多^[5,22]。这两个基因的产物通常只分泌到细胞的周质空间。但是,蔗糖酶的信号肽在一定条件下也可以满意地把蛋白质分泌到培养基。凝乳酶原基因在 SUC2 基因信号序列的指导下取得比较高的分泌性表达就是一个例子^[22]。

(四)影响酵母蛋白质分泌的其他因素

利用酵母系统进行分泌性表达,取得比较好结果的一般都是分子量较小的蛋白质。分子量较大的蛋白质似乎较难通过分泌途径。Emr^[7]等人发现把 *E.coli* 的 β -半乳糖苷酶和蔗糖酶的信号肽融合,产物只能进入内质网,不能分泌到周质空间。但一些事实说明,分子量为 40kd 到 70kd 的产物还是可以分泌的。例如: Rothstein 等^[18]和 Tubb 等^[23]分别克隆了小麦的 α -淀粉酶基因和 *S.diastaticus* 酵母的 α -1,4 葡萄糖淀粉酶基因。尽管它们的产物分别为 44kd 和 60kd,它们都能被酵母细胞所分泌。酵母本身的一个基因 MEL1 是编码 α -半乳糖苷酶的,这个酶的分子量达到 50kd,但是它在酵母培养基中的分泌量可达总分泌蛋白的 40%^[11]。特别有兴趣的是:当利用天然的启动子和信号肽来表达克隆在酵母细胞中的泡盛曲霉 (*Asperigillus awamori*) 葡萄糖淀粉酶基因时,尽管产物分子量高达 70—74kd,但分泌量可达 90% 以上^[14]。因此蛋白质被分泌的多少不仅和蛋白质分子大小有关,很可能还受蛋白质本质的性质的影响。这是值得深入研究的问题。

宿主细胞的突变可以阻断蛋白质的分泌,也有可以提高蛋白质分泌作用的突变。Bessey 等^[1]选到一株能提高杀死毒素分泌达 20 倍之多的突变体 UK-35。

Emr^[8]也分离到分泌突变体 Vpt3 和 Vpt5。Smith 等^[22]利用分泌凝乳蛋白酶原的菌落能在培养基表面留下一个痕迹的特性,用诱变剂 EMS 从 120000 个诱变菌落中得到 39 株超级分泌菌株 (Supersecreting strain)。这些菌株并没有增加对凝乳酶原的表达量,但是其中一些可以提高分泌能力达 8—10 倍,最强的两株称为 SSC1 和 SSC2。利用这种菌株,以蔗糖酶信号肽指导的凝乳酶原基因分泌量可达到 20mg/L。相当于细胞产生的凝乳酶原的 80—85%。用其中一株超级分泌菌株 SSC1-1 来分泌表达牛生长激素,分泌量至少提高 10 倍。

酵母对外源基因表达产物的分泌还同载体的存在形式有关,如果同样的载体结构被整合进染色体,可以得到比较高的分泌量。如果把四个拷贝带凝乳酶原基因的分泌载体 DNA 整合进酵母基因组,其分泌量和在游离状态时相比可提高 5—10 倍^[22]。这种分泌量提高的机制还不清楚。在实际上,这种组建方式是很有意义的。因为整合的基因十分稳定,在培养时可以不必要再用选择性培养基。

酵母对外源蛋白质的分泌还受细胞的生理条件的影响。Brake 等^[3]在利用 α -因子信号肽对人表皮生产因子 (hEGF) 基因进行分泌表达时,当培养物达到静止生长期后的 24h,分泌量达到 99% 以上,差不多占细胞总蛋白的 7%。而在活跃的生长期,大部分 (67% 以上) 仍留在细胞内。另外从经验知道,采用丰富培养基一般可以得到比较高的分泌水平。

总之,酵母细胞对外源蛋白质的分泌是遗传工程实践上一个十分有意义的课题。目前在这个方面的研究工作虽然已取得了一些进展,找到了一些提高分泌量的途径。但也引出了许多值得研究的理论问

题。例如: 不同外源蛋白质在分泌行为上为什么不同? 外源蛋白质的分泌过程和蛋白质的翻译后加工过程究竟有什么关系? 细胞分泌突变体的本质如何? 弄清这些问题将能进一步提高酵母对外源蛋白质的分泌表达量。

参 考 文 献

- [1] Bessey, H. et al., *Current Genetics*, 7:449—456, 1984.
- [2] Bitter, G.A. et al., *Proc. Natl. Acad. Sci. USA*, 81:5330—5334, 1984.
- [3] Brake, A.J. et al., *Proc. Natl. Acad. Sci. USA*, 81:4642—4646, 1984.
- [4] Cabezon, T. et al., *Proc. Natl. Acad. Sci. USA*, 81:6594—6598, 1984.
- [5] Chang, C.N. et al., Abstracts of papers presented at the 1983 Meeting on the Molecular Biology of Yeast, Cold Spring Harbor Lab.
- [6] Edens, L. and Van der Wel, H., *Trends in Biotechnology*, 3(3):61—64, 1985.
- [7] Emr, S.D. et al., *Molecular and Cellular Biology*, 4:2347—2355, 1984.
- [8] Emr, S.D. et al., In Protein Transport and Secretion, Gething, M-J, Ed, Cold Spring Harbor Lab., New York, p.190, 1985.
- [9] Hitzeman, R.A. et al., *Science*, 219:620—635, 1983.
- [10] Kurjan, J. and Herskowitz, I., *Cell*, 30:933—943, 1982.
- [11] Liljestrom, P.L., *Nucleic Acids Research*, 13:7257—7268, 1985.
- [12] Miyajima, A. et al., *Gene*, 37:155—161, 1985.
- [13] Mellor, J. et al., *Gene*, 24:1—14, 1983.
- [14] Numberg, J.H. et al., Glucoamylase cDNA, International Application Number: PCT/US 84/00122, International Publication Number: W084/02921.
- [15] Pennica, D. et al., *Nature (London)*, 301:214, 1983.
- [16] Roggenkamp, R. et al., *Journal of Biological Chemistry*, 260:1508—1512, 1985.
- [17] Roggenkamp, R. et al., *Proc. Natl. Acad. Sci. USA*, 78:4466—4470, 1981.
- [18] Rothstein, S.T. et al., *Nature*, 308:662—665, 1984.
- [19] Schekman, R. and Novick, P., In The Molecular Biology of the Yeast *Saccharomyces*, Vol. 2: Metabolism and gene expression, (J. Strathern, E.W. Jones and J.R. Broach Eds.), pp. 361—398, 1982.
- [20] Simons, G. et al., *Gene*, 28:55—64, 1984.
- [21] Singh, A. et al., *Nucleic Acids Research*, 12:8927—8938, 1984.
- [22] Smith, R.A. et al., *Science*, 229:1219—1224, 1985.
- [23] Tabb, R.S. et al., In Gene expression in yeast, Foundation for Biochemical and Industrial Fermentation Research (M. Korhola and E. Vaisanen, Eds.) Vol. 1, pp. 229—231, 1983.
- [24] Wood, C.R. et al., *Nature*, 314:446—449, 1985.