

地衣芽孢杆菌 2709 碱性蛋白酶基因的克隆、 序列测定及其表达

洪 扬 雷 虹 赵玉风 韩玉珉 申同健

(中国科学院生物物理研究所二室, 北京 100101)

摘 要 以克隆的地衣芽孢杆菌 2709 碱性蛋白酶编码序列的 PCR 扩增片段为探针, 通过原位杂交从 2709 基因文库中筛选出两个含有完整的 2709 碱性蛋白酶基因的阳性克隆: pSC1 和 pSC7。对 pSC7 中的插入片段构建若干亚克隆后测定了其全部 DNA 序列, 结果显示该插入片段含 2709 碱性蛋白酶及其信号肽与导肽 (Pro-peptide) 在内的全部编码序列 (1140 碱基对) 及长度分别为 299 和 832 碱基对的上、下游序列, 该序列同 M. Jacobs 等克隆的地衣芽孢杆菌 NCIB 6816 的 Subtilisin Carlsberg 基因序列显示了极高的同源性。通过枯草杆菌-大肠杆菌穿梭质粒 pBE2 将克隆的 2709 碱性蛋白酶基因转入到蛋白酶缺陷型的枯草芽孢杆菌 DB104 中, 结果表明 2709 碱性蛋白酶基因在枯草芽孢杆菌中得到了明显的表达。

关键词 地衣芽孢杆菌, 碱性蛋白酶基因, 克隆, DNA 序列分析, 表达

枯草杆菌蛋白酶的蛋白质工程是现今蛋白质工程研究中一个引人注目的领域⁽¹⁻³⁾。我们已通过 PCR 技术对工业用地衣芽孢杆菌 2709 碱性蛋白酶的编码序列作了体外扩增和全序列分析, 肯定了其分泌的碱性蛋白酶是典型的 Subtilisin Carlsberg 类⁽³⁾。本文报道通过原位杂交从 2709 菌株的基因文库中克隆到完整的碱性蛋白酶基因, 并测定其全部 DNA 序列, 还进一步完成了 2709 碱性蛋白酶基因在枯草芽孢杆菌中的表达。

1 材料与方 法

1.1 酶和试剂

限制酶、T4 DNA 连接酶、Taq DNA 聚合酶等均为 New England Bio-labs 产品。ATP 和 4 种脱氧核糖核苷酸为 Pharmacia 产品。异丙基硫代 β -D 半乳糖 (IPTG) 和 5-溴-4-氯-3-吡啶- β -D 半乳糖苷 (X-gal) 为 Sigma 公司产品。同位素 α -³²P, -dATP 和 α -³⁵S-dATP 为 DuPont 公司产品。M13/pUC 正反向测序引物购自 Pharmacia 和 Boehringer Mannheim。蛋白酶测活底物 succinyl-L-Ala-L-Ala-L-Pro-L-Phe-p-nitroanilide 和抑制剂 PMSF 购自 Sigma 公司。

1.2 菌株和质粒

地衣芽孢杆菌 2709 购自中国科学技术大学生物系, 大肠杆菌 DH 5 α MCR 购自 BRL 公司, JM109 和 GM2163 为本实验室保存, 枯草芽孢杆菌 DB104 由中国科学院遗传所李心治先生馈赠。pUC18 和 pUB110 由本实验室保存, 枯草芽孢杆菌-大肠杆菌穿梭质粒 pBE2 由郭兴华先生馈赠⁽⁴⁾。pUC18-A1 和 pUC18-A2 为插入有 2709 碱性蛋白酶成熟蛋

白编码序列之 PCR 扩增片段的大肠杆菌重组质粒, 由本实验室构建^[3]。

1.3 培养基

芽孢杆菌和大肠杆菌的普通培养均采用 LB 培养基, 视需要分别加入 Agar (1.5%), X-gal (300 μ g/ml)、IPTG (80 μ g/ml)、氨苄青霉素 (100 μ g/ml)、卡那霉素 (50 μ g/ml) 和脱脂奶粉 (1%, w/v; 2709 菌株培养用)。DM3 再生培养基的配制参见文献[5]。

1.4 2709 染色体 DNA 的提取

大体按 Marmur 方法^[6], 略有修改。

1.5 DNA 重组操作

参见文献[7]。

1.6 枯草芽孢杆菌的原生质体转化及质粒提取

原生质体转化按文献[5]进行。枯草杆菌中质粒的制备, 参见文献[7]。

1.7 DNA 序列测定

采用双脱氧法。测序样品制备按 Pharmacia 的 T7 Sequencing Kit 和 Deaza Sequencing Kit 说明书进行, 标记同位素为 α -³⁵S-dATP。测序凝胶厚度为 0.25 或 0.4mm, 长度为 55—80cm, 电泳电压为 50V/cm。

1.8 蛋白酶活力测定

参见文献[8]。

2 结果与讨论

2.1 地衣芽孢杆菌 2709 基因文库的构建

按 Shotgun 策略采用末端半补齐技术构建基因文库^[9,10], 共获约 3×10^4 个转化子, 其中 70% 以上是重组子。这不仅满足完全文库构建的条件 (不少于约 5×10^3 个重组子), 而且大大减轻了对文库进行筛选的工作量。

2.2 2709 碱性蛋白酶基因的筛选

从 pUC18-A2 中用 BamHI 切出 2709 碱性蛋白酶编码序列的 PCR 扩增片段 (约 844bp), 经放射性标记后作为对 2709 基因文库进行原位杂交的探针。参考 Wells 等人^[8]的工作, 我们把筛选过程分为三步: (a) 基因文库的大量原位杂交筛选: 利用上述探针, 用大量原位杂交的方法分两批筛选了所构建的 2709 基因文库, 每批约 1.5×10^4 个转化子。这两批杂交均分别获得了 8 个阳性克隆; (b) 阳性克隆的定位与两次杂交: 从基因文库中将被认为是阳性的菌落及其周围若干菌落一并挑出, 按小量原位杂交的方法进行第二轮筛选。在高严谨洗脱条件下, 有 4 个表现为阴性而遭淘汰; (c) 阳性克隆的单克隆杂交与部分序列分析: 将 12 个阳性克隆各自分离单菌落后, 分别挑选 8—10 个单菌落再进行一轮原位杂交。从中选出 12 个呈杂交阳性的单克隆, 通过初步的酶切图谱分析, 推测其中的 pSC1 和 pSC7 含有完整的 2709 碱性蛋白酶基因。对 pSC1 和 pSC7 作了部分序列测定, 结果表明它们不仅均含完整的 2709 基因, 而且还具有相同的插入方向和基因 5' 端, pSC7 的物理图谱见图 1。质粒 pSC1 与 pSC7 的区别仅在于前者插入片段的 3' 端较后者长 900bp 左右。

2.3 地衣芽孢杆菌 2709 碱性蛋白酶基因的全序列分析

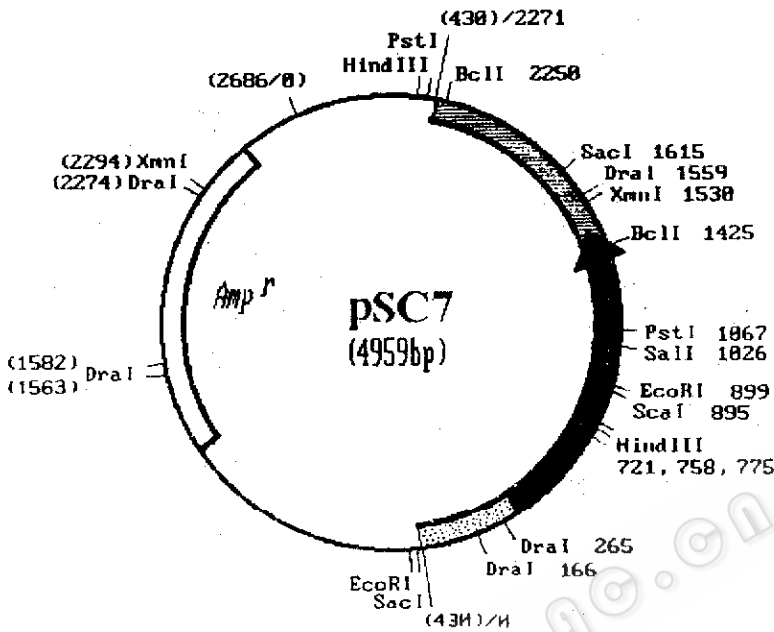


图 1 pSC7 物理图谱

Fig. 1 Physical map of plasmid pSC7

■ The coding sequence of the subtilisin Carlsberg gene from *B. licheniformis* 2709

▨ The flanking sequence of the subtilisin Carlsberg gene from *B. licheniformis* 2709

我们挑选了插入片段相对较短的 pSC7 作为测定 2709 基因全序列的材料，其序列测

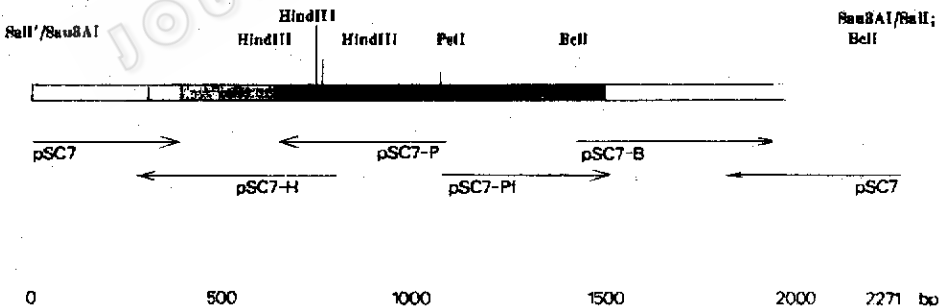


图 2 2709 碱性蛋白酶基因的测序策略

Fig. 2 Sequencing strategy for the subtilisin Carlsberg gene from *B. licheniformis* 2709

▨ Flanking sequences ▨ Signal-peptide coding sequence

▨ Pro-peptide coding sequence ■ Mature subtilisin coding sequence

定策略如下：按图 2 设计，pSC7 中的插入片段被划分成 6 个大致等长的测序区域，并为此构建了 4 个亚克隆，它们的构建过程大致为：

- (a) pSC7-H：将 pSC7 用 Hind III 酶切除去 ~1.5kb 降解片段后，再经自环化即为亚克隆 pSC7-H；
- (b) pSC7-P：将 pSC7 用 Pst I 酶切除去 ~1.1kb 片段后再自环化即为亚克隆 pSC7-P；
- (c) pSC7-Pf：将 pSC7 的 Pst I 酶切片段 (~1.1kb) 插入到 pUC18 的 PstI

位点中,即构建成亚克隆 pSC7-Pf; (d) pSC7-B: 由于 Bcl I 切点受大肠杆菌 dam 基因产物甲基化的影响,为此先将 pSC7 转入到基因型为 *dam-dcm* 的 GM2163 菌株中,提取无甲基化碱基的 pSC7 后,再用 Bcl I 切出 2709 基因 3'端约 800bp 的片段,插入到 pUC18 的 BamHI 位点中,即构成亚克隆 pSC7-B; pSC7 自身则被用于测定 2709 基因 5'和 3'两端的序列。

2.4 2709 碱性蛋白酶基因的序列测定结果

pSC7 中 2709 碱性蛋白酶基因的序列测定结果见图 3。其插入片段的长度为 2271bp (图 3 中仅显示至下游 Sac I 切点前 1615 bp 的序列),包括了 2709 碱性蛋白酶基因 299bp 的上游序列,1140bp 的结构基因序列和 832 bp 的下游序列,借助有关计算机软件,将 2709 基因序列同 M. Jacobs 等^[11]克隆的地衣芽孢杆菌 NCIB 6818 序列以及其它枯草芽孢杆菌蛋白酶基因序列^[8,12]作了若干比较分析,我们发现:(a)在基因 5'端上游序列可比较的范围内,NCIB 6816 同 2709 基因是完全一致的。将 2709 基因的上游序列同 subtilisin E 的上游序列作仔细的同源性比较分析后发现,尽管二者有相当的差异,但在过渡态调节因子 AbrB 结合序列(约 80bp)^[13]的范围内却表现出明显的同源性,说明其调控序列具有较大的保守性,这提示 AbrB 基因在芽孢杆菌的碱性蛋白酶的表达调控中占有重要地位^[14]。同源性分析显示出 2709 基因的启动子最有可能位于 189—194 位碱基(相当于-35)和 213—218 位碱基(相当于-10)处,这与新近有关 NCIB 6816 基因启动子序列鉴定的实验结果是一致的。(b)2709 碱性蛋白酶的结构基因全长 1140 bp,编码了包括信号肽、导肽及成熟的蛋白酶在内的共 379 个氨基酸残基的多肽,其中信号肽与导肽部分无论在氨基酸序列还是在 DNA 序列上均与 NCIB 6816 基因完全一致。但在编码成熟蛋白酶的部分两者有相当的差异,有关详细的比较分析可参见文献[3]。值得提及的是,通过对 2709 基因的全序列测定,可以确定在我们早期对 2709 基因所做的 PCR 扩增工作

表 1 2709 碱性蛋白酶基因在枯草芽孢杆菌 DB104 中的表达

Table 1 Expression of 2709 alkaline protease gene in *B. Subtilis* DB104

Strain (plasmid)	Protease activity ($\Delta OD_{412nm}/min/ml$)
DB104 (pBE2)	1.2×10^{-3}
DB104	0.17
(pBE2-SC7) +PMSF	0 *
+EDTA	0.16 * *
<i>B. licheniformis</i> 2709	10.08

Note: * $5\mu l$ 100mmol/L PMSF (in ethanol) was added to 500 μl culture supernatant and incubated for 10 min at room temperature

* * : $10\mu l$ 1500mmol/L EDTA (pH8.0) was added to 500 μl culture supernatant and incubated for 10 min at room temperature

中^[3],所测两条 PCR 扩增片段序列中的一个(pUC18-A2),在 562 位(2709 基因序列的 1075 位)有一个 PCR 扩增所导致的碱基错配(C→T),除此以外,PCR 所测序列同 2709 基因全序列的测定结果是完全吻合的。这说明对一个不长的序列进行 PCR 扩增,只要 PCR 扩增条件合适,其扩增结果准确性还是比较高的;(c)在下游序列部分,2709 基因同 NCIB6816 基因在可比较的范围内仍是完全一致的,包含一个明显的不依赖于 ρ 因子的转录终止信号(图 3 中的波浪线部分所示),在此信号下游约 120bp 处有一唯一的 Sac I 切点,为 2709 基因的进一步基因工程操作提供了

很好的条件。

2.5 2709 碱性蛋白酶基因在枯草芽孢杆菌 DB104 中的表达

参照序列分析的结果, 我们首先采用 Sac I 和 BamH I 将 2709 基因从 pSC7 中完整的分离出来, 该片段全长约 1.6kb, 包括了 2709 碱性蛋白酶基因完整的结构基因 (1140bp)、上游序列 299bp 以及含转录终止信号在内的下游序列 180bp。将此片段插入到 pBE2 质粒的 BamH I 和 SacI 切点间, 构建成表达质粒 pBE2-SC7。根据对序列测定结果的分析及 Suchlein 等的工作^[15], 我们期望 2709 基因能在自身启动子的指导下在枯草芽孢杆菌中表达并分泌到胞外。

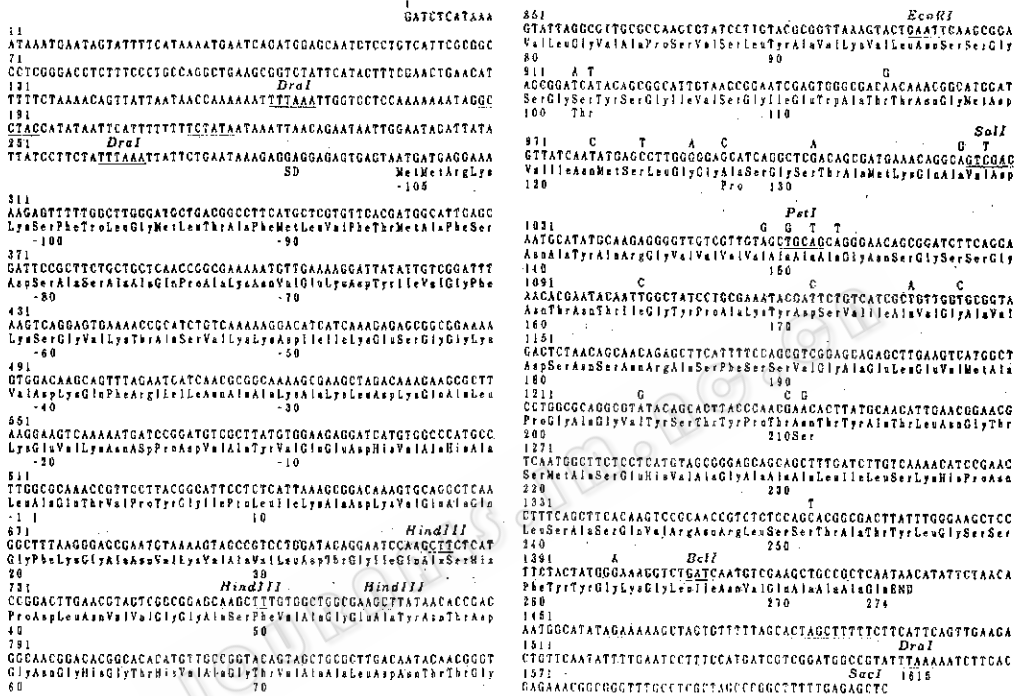


图 3 2709 碱性蛋白酶基因的 DNA 序列及相应的氨基酸序列

Fig. 3 Nucleotide and amino acid sequence of the subtilisin Carlsberg gene from *B. licheniformis* 2709

Note: The continuous sequence is that of subtilisin Carlsberg gene from *B. licheniformis* 2709. The nucleotides, amino acids that are different compare to the published sequence of subtilisin Carlsberg gene from *B. licheniformis* NCIB 6816 are shown above or below the continuous sequence

~~~~~ The possible promoters and putative transcription termination sequence

Putative SD sequence

将 pBE2-SC7 与 pBE2 分别通过原生质体转化导入到蛋白酶双缺陷型的枯草芽孢杆菌 DB104 中, 挑选单个转化子于含卡那霉素 (50μg/ml) 的 LB 液体培养基中, 37℃ 振荡培养 24 小时, 将菌液离心后取上清测定其碱性蛋白酶活力, 结果见表 1。从表中数据可知, DB104 (pBE2-SC7) 的培养液上清的蛋白酶活力, 较对照菌株 DB104 (pBE2) 约高 100 倍以上, 并受丝氨酸蛋白酶抑制剂 PMSF 的专一性抑制。与此同时, DB104 (pBE2-SC7) 菌落在含 1% 脱脂奶粉的 LB 培养基上周围还可以明显观察到由于蛋白酶分泌而形成的透明圈, 而在 DB104 (pBE2) 菌斑周围则仅有极微弱的透明圈出现。同 2709 原始菌株相比, 尽管 pBE2-SC7 在 DB104 中可能有较高的拷贝数, 但 2709 碱性蛋白酶基因在 DB104 中的表达水平仍然是比较低的。结合对 2709 基因上游序列分析比较的结果, 可以

认为这主要是因为 2709 的上游序列与枯草芽孢杆菌差异较大所致。

致谢 中国科学院微生物所郭兴华研究员提供穿梭质粒 pBE2, 特此致谢。

### 参 考 文 献

- [1] Wells J A *et al.* TIBS, 1988, 13: 291.
- [2] 赵云德等. 生物工程进展, 1991, 6: 16.
- [3] 雷 虹等. 生物化学杂志, 1993, 9: 411.
- [4] 郭兴华等. 生物工程学报, 1991, 7: 224.
- [5] Chang L *et al.* Mol Gen Genet, 1979, 168: 111.
- [6] Marmur J. J Mol Biol, 1961, 3: 208.
- [7] Sambrook J. *et al.* Molecular Cloning (laboratory manual second edition), Cold Spring Harbor Laboratory Press, 1989.
- [8] Wells J A *et al.* Nucleic Acids Res. 1983, 11: 7811.
- [9] Hung M C *et al.* Nucleic Acids Res. 1984, 12: 1863.
- [10] 洪 杨等. 生物化学与生物物理进展, 1993, 20: 467.
- [11] Jacobs M. Nucleic Acids Res. 1985, 13: 8913.
- [12] Stahl S L *et al.* J Bacteriol, 1984, 158: 411.
- [13] Strauch M A *et al.* The EMBO J. 1990, 8: 1801.
- [14] Burbulys D *et al.* Gell, 1991, 64: 545.
- [15] Suhlein R *et al.* Mol Gen Genet, 1991, 227: 137.

## Cloning, Sequencing and Expression of the Alkaline Protease Gene from *Bacillus licheniformis* 2709

Hong Yang Lei Hong Zhao Yufeng Han Yumin Shen Tongjian

(Institute of Biophysics, Academia Sinica, Beijing, 100101)

**Abstract** Using the PCR amplified partial coding sequence of the alkaline protease (subtilisin) from *Bacillus licheniformis* 2709 as the probe of *in situ* hybridization, several positive colonies were identified from the 2709 genomic library. Two of them, pSC1 and pSC7, were shown having the intact 2709 gene by physical mapping and partial sequencing. The complete sequence of the insert of pSC7 plasmid has 2271 base pairs including 299bp 5' flanking sequence, 832bp 3' flanking sequence and 1140bp structural sequence. The 2709 subtilisin sequence obtained here is highly homologous to the subtilisin from *B. licheniformis* NCIB 6816. The cloned gene was introduced into the protease-deficient *Bacillus subtilis* DB104 and protease activity assay demonstrated that a successful expression of 2709 subtilisin in DB104 was achieved.

**Key words** *Bacillus licheniformis*, alkaline protease, subtilisin, cloning, sequencing, expression